Faculty of Philosophy,
Theology and Religious Studies

Master's Thesis

Compatibilism and Actual Miracles

Davide De Mola

examiners:

Prof. N.P. Landsman

Prof. M.V.P. Slors

Radboud University

Contents

1 Introduction **1**

2 The Problem of Free Will **3**
2.1 The landscape of the problem 3
2.2 From free will to the Free Will Thesis 7
2.3 Determinism 10
2.4 Lewis and van Inwagen 12

3 The Consequence Argument **13**
3.1 Timeline of debate 13
3.2 The core problem 15
3.3 Van Inwagen's First Formal Argument 16
3.4 Lewis's reply: Weak and Strong abilities 17
3.5 Arguments for Local Miracle Compatibilism 19
3.5.1 From counterfactuals to miracles 19
3.5.2 Premise *5* and the principle of disglomeration 22
3.6 Van Inwagen's Third Formal Argument 25
3.7 Moral of Chapter 3 26

4 New Challenges to LMC **27**
4.1 The Free Will Theorem 27
4.2 Landsman on the Free Will Theorem 29
4.2.1 Interpreting the *Determinism* premise 31
4.2.2 Interpreting the *Freedom* premise 35
4.2.3 *Freedom* interpreted 39
4.3 Moral of the Free Will Theorem 44
4.4 Dorr Against Counterfactual Miracles 44
4.5 Dorr's First Argument 46
4.5.1 Dorr's argument from Statistical Mechanics 48
4.5.2 Macrohistories, macrostates and S2 49
4.5.3 The Independence Conjecture 50
4.5.4 From independent probabilities to existence 56
4.6 Moral of Dorr's Arguments 57

5 Conclusions **59**
5.1 Compatibilism without Actual Miracles 59

6 Bibliography **61**

7 End Matter **75**
7.1 Acknowledgments 75
7.2 Certification of Ownership 76
7.3 Acronyms 77

1 Introduction

In this thesis I will assess two new arguments on physical grounds against Lewis's Local Miracle Compatibilism. I will argue neither is decisive, but for different reasons and with different *caveats*. Actual miracles, that is, events not in keeping with our laws of nature, play a role in both reasons.

The former argument, put forward by Landsman, must rely on a principle of recombination of possibilities which Lewis can plausibly deny. However it does succeed in bringing forth a new problem with Lewis's position.

The latter argument, put forward by Dorr, is flawed due to its reliance on a physical conjecture put forward elsewhere by Wallace, which I also critically assess.

These are the narrowly defined topics of this work. However, I discuss both these arguments in the wider context of the free will debate. In particular, I examine how they affect the dialectical stalemate between van Inwagen and Lewis over the Consequence Argument for incompatibilism.

I will suggest these two arguments and their physical framework offer a new perspective on Lewis's distinctive version of compatibilism which I contrast with the 'libertarian compatibilism' recently defended by List. My tentative conclusion is that List's ideas retain the attractive features of Lewis stance without the need for miracles. It seems like a fruitful line to pursue for a compatibilism-inclined physicalist.

Chapter 2

The Problem of Free Will

The problem of free will stems from a familiar question for physicalists like myself, which Jackson has dubbed the location problem. Physics posits a very short list of fundamental entities and properties: the *physical* features of the world. We believe they truly are features of the world because our best theories mention them explicitly. We also believe that all other features of the world supervene on, or even reduce to, the physical ones. Thus we are faced with a choice. Given a putative feature of the world which does not appear on the list, we must do as Jackson says and "either eliminate or locate" it (2000, p. 5).

One such putative feature of the world is particularly close to our hearts. What is *our* place in this grand scheme? What is the place of the *wilful agents* we usually take ourselves to be? In particular, what is the place of our pre-philosophical notion of free will (if there even is such a thing) in the physical universe? Although just a corner of the general "location problem" we physicalists face, it is one whose urgency has exercised philosophers since Democritus.

2.1 The landscape of the problem

The problem of free will in bare outline is this: we *cannot* locate it and we do not *want* to eliminate it either. In more detail, the problem of free will may be stated as a trilemma following van Inwagen (2008, p. 327):

T1 There are seemingly unanswerable arguments that (if they are indeed unanswerable) demonstrate that free will is incompatible with determinism.

T2 There are seemingly unanswerable arguments that demonstrate that free will is incompatible with indeterminism.

T3 There are seemingly unanswerable arguments that demonstrate that the existence of moral responsibility entails the existence of free will.

I will tighten the trilemma shortly by defining its terms, but I can already highlight its impact on the location problem. If T1 and T2 are true together the prospects of locating free will are dim. We face the law of excluded middle. Either determinism or indeterminism holds and neither is compossible with

free will. It follows that, if we accept this potted argument, we cannot hope to locate free will. It *cannot* exist as a matter or logical necessity, let alone be located in a physicalist worldview.

This would not be the first time physics forces us to revise cherished beliefs, perhaps our belief in the existence of free will is among them. However any revision comes at a cost and for free will the cost is prohibitively high according to T3. It would not be a revision, but a wholesale rewrite of our self-image: this is the sense in which *we* do not *want* to eliminate it. T3 grounds this reluctance to eliminate free will in its alleged connection to moral responsibility. Suppose for a moment we accept T3 as it stands. Then the extent to which we accept our ordinary moral discourse *at face value*, particularly our practice of attributing moral responsibility to certain actions, is the extent to which we do not *want* to eliminate free will.[1]

This may not be very much, with moral anti-realists of various stripes leading the charge. Given they have already shouldered the philosophical cost of not taking our moral discourse at face value, they could now reap their reward by flatly rejecting T3 and escaping the trilemma.

Alternatively some believe the cost of moral anti-realism is still too high and seek to undermine T3. Moral responsibility may well exist, roughly in the way we speak about it, but it is not connected to free will in the way T3 asserts, or at all perhaps. Thus to doubt the latter does not endanger the former. Indeed so far we only have van Inwagen's word that T3 is grounded in unanswerable arguments. I for one, harbour some doubts that the widespread acceptance (Vihvelin 2015, §1) of the connection between free will and moral responsibility rests on unanswerable arguments rather than more complex, more defeasible considerations. My doubts stem from two sources.

Firstly, from Hume onwards, empirically minded philosophers have been suspicious of overly strong connections between moral claims and factual ones. Suppose moral responsibility is itself grounded in a judgement about how things *ought* to be in relation to an agent's part in bringing them about, whereas free will is a claim about how things *are* given the laws. Then we should be suspicious of any unanswerable argument to underwrite T3. Any such argument must escape Hume's guillotine: you cannot derive an 'ought' from an 'is'. However, if we insist on a reading of 'cannot derive' which includes denying any stipulative connection we risk rejoining the moral anti-realist camp as eliminativists about moral responsibility. To wit: we uphold physicalism; deny it's possible for moral responsibility to be connected to physical facts in any conceivable way and are thus forced to eliminate it.[2] I have already discussed this line of argument above, so I will pursue it no further. On a weaker reading of 'cannot derive' we allow 'derivations' containing stipulative connections. The Kantian principle 'ought implies can' would be a salient example of the kind of constraints we might wish to impose on such connections:

[1] Cf. (4) in van Inwagen (2015, p. 21).

[2] I assume none of the laws of physics prescribe what is right or wrong.

> When we say absolutely of ourselves or others, "I ought to do so and so" or "you ought to", we imply, I think, very often that the thing in question is a thing which we could do, if we chose; though of course it may often be a thing which it is very difficult to choose to do. Thus it is clear that I cannot say of anyone that he ought to do a certain thing, if it is a thing *which it is physically impossible for him to do*, however desirable it may be that thing should be done.
>
> (Moore 2005, p. 140, my italics)

Perhaps these constraints are so desirable, so universally assumed as to constitute *de facto* unassailable assumptions. I doubt this is the case, which brings me to my second worry.

Secondly, I am unsure whether we may licitly refer to the moral discourse above as our *ordinary* one. Study after study in experimental philosophy shows philosophers' intuitions diverge significantly from those of non-philosophical folk when attributing moral responsibility (Nichols 2011; Sarkissian et al. 2010). Thus to justify T3 by claiming this is how we commonly talk of these issues is suspect. Perhaps the weaker claim that philosophers' discourse is a systematization of the ordinary one—what ordinary folk would say, if they thought about it as carefully as we do—may be enough. But now physicalists are faced with two tasks: one is to locate the free will-of-the-philosophers; the other is to show it is the best of imperfect deservers of the name free will-of-the-street. The rivers of ink spent on the connection between free will and moral responsibility could provide a starting point for the latter task, *pace* Schlick (1939, p. 143). Nonetheless we should also be prepared to leave the armchair and talk to moral psychologists to keep our theorizing honest.

All in all, I think T3 is the premise that provides the physicalist with the most room to manoeuvre in order to escape the trilemma. But there is a fine line to tread: on the one hand we want to keep our two tasks manageable and distinct. On the other, we want to ensure the free will we end up arguing about does not lose its relevance to everyday discourse, including moral discourse.

This tension can be seen in van Inwagen's (2008, fn. 3) advice *not* to define free will "as whatever sort of freedom is required for moral responsibility". Whilst this definition would secure the truth of T3, it would put the metaphysical cart before the normative horse. It is good advice because there is no *a priori* guarantee such a thing exists—so we should not glibly define it into existence. It is good advice for the sake of sanity—we do not want to talk past each other because we implicitly hold different ideas about moral responsibility.

It is especially good advice to the physicalist, for whom normative principles and moral responsibility in particular, are a yet un-located feature of the world. But it becomes bad advice if we lose sight of what makes free will a 'relevant' problem in the first place.

Broadly speaking there are two possibilities. Either we remain neutral about T3 and proceed under the assumption it *may* be true—it is possible the location problem for free will is 'coupled' to the moral responsibility one. But we reserve the right to sort out the details later. Or we reject T3 directly, as Fischer

and his fellow semi-compatibilists do—although for very different reasons to the moral anti-realists above. Semi-compatibilists say moral responsibility exists *and* that its existence does not entail the existence of free will.[3] I think Fischer's stance is partly motivated by the belief that, on the balance of evidence, the prospects of the former stance are hopeless as it inevitably results in endless stalemates (2012a, p. 9).

According to Fischer, one such stalemate is the debate between van Inwagen and Lewis (Vihvelin 1998, p. 414) over the Consequence Argument. In this essay I wish to re-examine this stalemate in the light of new evidence in the form of two recent arguments from physical grounds (Dorr 2016; Landsman 2016) Thus I assume the trilemma stands for the definition of free will I shall adopt and that the philosophical action lies with the Consequence Argument. So I also adopt the former stance on T3. At worst, Fischer was right all along and we will end up reconfirming his findings.

I note that Lewis, also an avowed physicalist (Lewis 1983b, p. xi), did not feel obliged to resolve the status of T3 in order to defend his denial of T1 (i.e. his compatibilism). And his opponent van Inwagen (1983, Ch. 5), has defended T3 at length resulting in an endless stalemate with Fischer over the the Principle of Alternative Possibilities and related attempts to establish T3.[4] Thus, my stance is partly motivated by the belief that this latter stalemate is even more intractable.

Pragmatically then, the former stance seems like a good starting point and there is no reason to suppose we cannot make some progress on free will independent of T3 (recall: we wish to keep the task manageable!).
This is also the course advocated by Earman:

> To state the issues in as neutral a way as possible, let us drop for the moment questions about moral responsibility, guilt, and punishment and begin instead by asking questions like: How are the actions of man different from those of a sunflower as it turns to face the sun? If determinism is true, aren't all of our actions merely complicated cases of tropisms, forced motions produced by circumstances beyond our control?
> [...] it is not just the Libertarians who feel the crunch of determinism but anyone who wants to accord man a special place in nature on the grounds that, in contrast to inanimate objects and the lower life forms, we enjoy an autonomy in that what we do is up to us.
>
> (Earman 1986, p. 239, 241)

This amounts to a special difficulty for the physicalist:

T3* There are seemingly unanswerable arguments for the physicalist that (if they are indeed unanswerable) demonstrate that determinism entails human actions are tropisms.

[3] "Thus, an agent can legitimately be held morally responsible for his behavior, even though he lacks regulative control (or freedom to choose and do otherwise)" (Fischer 2012b, p. 120).

[4] The Principle of Alternative Possibilities says that 'A agent is morally responsible for an action only if they could have done otherwise'. Following the publication of (Frankfurt 1969) the dialectic has come to worryingly resemble that generated by Gettier-cases in epistemology.

I will set aside the heavy handed reply that physicalists *already* believe that humans and sunflowers are in the same garden insofar as they are both physical entities. Rather Earman points out how human actions "are mediated by mental states" and in that sense already different from the actions of sunflowers.

I would add a third denizen to Earman's garden: fundamental fields.[5] I think we should ask ourselves how mental states relate to bodily states and actions (as Earman does) *and* how bodily states and actions relate to fields. *Prima facie* they all occupy different levels so we should clarify how deterministic behaviour (tropisms) arises at each level and whether it all goes back to the determinism of the most fundamental description. This will be an overriding theme of Chapter 4.

Finally Earman's example highlights the fact that there is also a non-normative aspect to free will which is also 'important to us'. This is the lived-in feel of making choices and of agential autonomy.

I think we have danced at the edge of the problem long enough. I hope you have seen enough of the landscape ahead to be convinced the journey is worth making. Without further ado I will now define free will and determinism and discuss the panoply of philosophical theses they engender.

2.2 From free will to the Free Will Thesis

Alice is a normal human being (with a healthy brain and body) going about her normal day to day activities. From time to time she will need to make *decisions*, that is, she will need to choose amongst alternative courses of action, for example picking a movie to watch from her streaming subscription.[6] Hopefully we can agree this accurately reflects the way we *talk* about ourselves. Does she have free will?

In common parlance it is usually enough to define free will and closeby concepts by a *via negativa* approach. A free human is someone who is not a slave or a prisoner. A free vote is an act of voting which is not bound by the Whip. These definitions work tolerably well because they tell us something relevant in their native contexts. A freeman may not be bought and sold. A free vote will not cause the voter to be be thrown out of the party. We could cheerfully generalize along with the OED (2016) and say Alice acts freely as long as she acts "without restraint or restriction upon action or activity; without hindrance, inhibition, or interference". If the lexicographers have done their

[5] These fields would be the primitive constituents of everything in some hypothetical deterministic Theory of Everything.

[6] This strikes me as a paradigmatic case of deliberation-and-then-choice we *routinely* believe ourselves to engage in. Note that: i) external constraints remain *ceteris paribus* amongst the various options: cost, time, availability, etc. ii) the internal constraints *appear* to do most of the work: general taste in film, mood etc. iii) in some philosophical discussions of the sort 'I choose to raise/not raise my hand' it almost seems the agent is purposefully going out of her way to exercise this 'faculty of free will' we suppose her to have. I also agree with Earman (1986, p. 248) that sometimes it is the smaller, least-morally-momentous decisions in life where we later report *feeling* most free.

job, then this will be an accurate portrayal of how we talk about free will.

Let us follow this path further. We start by considering humdrum constraints on Alice's actions and then inquire into more and more wide-ranging factors that could prevent her from acting freely. As philosophers we end up worrying about constraints which are imposed not by fellow humans, but by the very laws of nature, and not by whips or chains, but by means that would make a conspiracy theorist proud. How can we tread this path to a definition, when potentially any factor could be conditioning her? We can hardly ask for the external world to be absent just so we may be sure it is not actually restricting her in any conceivable way. We must step off the via negativa and instead ask that, whatever else obtains, it should nonetheless fail to prevent our 'will' in 'its' intent. It should somehow be above the fray of coercion and compulsion.

The experimental philosopher Joshua Knobe employs the following useful metaphor for this way of understanding human action:

> Consider a royal court. The advisors and ministers each have an opportunity to advocate for a particular course of action. But it is not as though the advisors and ministers themselves make the final decision. Instead, there is another person in the court—the king or queen—who listens to all of the arguments, thinks them over, and then decides.
>
> The mind works in more or less the same way. Your mind might include various states and processes, but it would be a mistake to suggest that you yourself are just a collection of states and processes. On the contrary, you are a further thing—like the king or queen in the court—who can attend to the states and processes within your mind and then freely make a choice.
>
> When you do end up making a free choice, we might say that you made this choice 'on the basis of' some of your psychological states. But the connection here is always indirect. It is not as though your psychological states actually cause your behavior; you just freely decide what to do, and sometimes you end up deciding to act in a way that accords with them.
>
> (Knobe 2014, p. 69)

According to Knobe, this metaphor also captures the prevailing folksy intuition about how the mind works and how free choices are made. Just as the king's power is the ultimate arbiter of the land, subject only to wind and tide, so with the will during decision making: it is limited only by *external* impossibilities, but is otherwise unconstrained. Or to put it another way: overarching external factors fix the *possible* alternatives, but, *given* those alternatives, the will 'decides'. Thus 'Alice chose freely' iff, upon reflection, she can say she was in the same position as the king when deliberating. She weighed up the options, her mood, her preferences and then settled on her final choice without being coerced in any way.

Metaphors will only carry us so far. In particular we must not prejudice the question by defining free will in terms of choices. Using the word 'choice' implies the existence of alternatives *and* that Alice is somehow able to select amongst them. This brings me to the *philosophical* definition my two main contenders agree on. Compare:

> I have just put my hand down on my desk. That, let me claim was a free but predetermined act. *I was able to act otherwise*, for instance to raise my hand,
>
> (Lewis 1981, p. 113, my italics)

with

> When I say of a man that he "has free will" I mean that very often, if not always, when he has to choose between two or more mutually incompatible courses of action—that is, courses of action that it is impossible for him to carry out more than one of—each of these courses of action is such that he can, or *is able to*, or has it within his power to carry it out.
>
> (van Inwagen 1983, p. 8, my italics)

Note the subtle shift from the lived-in feel of the folk conception to the austere third person perspective of the two philosophical definitions. Undeniably something is lost in the passage from one to the other. Hopefully the philosophical definition captures, at the very least, a necessary condition for the richer folk conception to apply.

I am ready to give the definition of free will I will use for the remainder of this essay. Thus 'Alice acted freely' iff

FWT Alice was able at time t to raise her hand and thus watch a different movie to the one she in fact watched.

If we take FWT as a claim rather than a *definiens* we have the Free Will Thesis.

Of course we really should quantify more carefully. What we *really* would like to say is this:[7] *Whomever* is relevantly like us, is free, *whenever* they reasonably think they are. Who is relevant? When is reasonable?

Sunflowers are not free and neither are people with certain brain tumours,[8] but we should not exclude intelligent aliens or maybe the more advanced of our fellow *Animalia*. Regarding our own actions, the evil mastermind's heist was freely conceived, but not his dash when the alarms sounded, or my own seemingly free choice of ice-cream (I just *love* pistachio). In short, the search for a cut-and-dried criterion for who and when we wish to claim free will seems hopelessly messy. Typically philosophers just dodge the problem. Worry about securing free will at least once for at least one idealized agent! That is enough to decide T1.

True enough, but I cannot dispel the uneasiness this attitude leaves behind. Suppose they got exactly what they wished for, Alice was indeed free at time t to raise her hand. Good for her! What use is that to the average Joe? I just wish to point out this epistemic problem. Unless our messy everyday judgements are a good guide to what philosophers decide 'being able to do otherwise' means, it is all for naught. We may think we are free but not be so, and we may be free but not think so. This concern can only increase if arguments over 'being able

[7] Unless of course you are already utterly convinced of the truth of T1!

[8] See (Burns and Swerdlow 2003) for a particularly harrowing and clear cut case of orbitofrontal tumours affecting personality and agency.

to do otherwise' are ultimately decided with reference to fundamental physics. Concern noted, let us press on.

2.3 Determinism

This section will be a series of footnotes to Earman, whose magisterial book *A Primer on Determinism* (1986) is still the starting point for any serious study of this venerable doctrine. Thus:

> Letting $\mathcal{W}$ stand for the collection of all physically possible worlds, that is, worlds which satisfy the natural obtaining in the actual world, we can define the Laplacian variety of determinism as follows. The world $w \in \mathcal{W}$ is *Laplacian deterministic* just in case, for any $w' \in \mathcal{W}$, if w and w' agree at any time then they agree for all times. By assumption, the world-at-a-given-time is an invariantly meaningful notion and agreement of worlds at a time means agreement at that time on all relevant physical properties.
>
> (*ibid.*, p. 13)

I will adopt the Ptolemaic version of Earman's definition:

Det Determinism is the claim that our world belongs to a set $\mathcal{W}$ of Laplacian worlds: they obey the actual laws and they never branch *because* of that.[9]

This immediately puts determinism and free will on a collision course. Whereas we model an agent's act of choosing as a garden of forking paths, according to determinism it's one way or the miraculous way.

Hence the reference to the actual laws is crucial in two respects. Firstly, much of what is to follow will hang on whether a slight weakening of this requirement is philosophically viable. This will take us into the debate over the nature of lawhood and of possible worlds.[10] But I can already stipulate an actual miracle is an event which does not conform with the actual laws, i.e. the laws that obtain in the actual world.

Secondly, it embodies a certain kind of attitude to the question of determinism which I espouse wholeheartedly. Namely, the truth of determinism is very much a scientific question: in particular we should look at the laws of physics to find out whether (fundamental) determinism holds. And this is precisely what Earman goes on to do:

> Most of the putative laws of physics take the form of differential equations for which questions of determinism principally involve existence and uniqueness properties of solutions.
>
> (*ibid.*, p. 21)

[9] I implicitly assume that by appealing to possible worlds we ensure $\mathcal{W}$ is also not empty, nor a singleton. There are some further issues that might worry us which Earman (1986, Ch. 2, § 7) discusses which I have collapsed into my italicized 'because'.

[10] In particular my definition of Det might suggest determinism is making a strong ontological claim about the existence of other worlds, so that an appropriate set $\mathcal{W}$ exists for our own world to belong to. Lewis would have certainly favoured this reading, but this is by no means required. The standard understanding of possible worlds would entail Det is a claim about that type of *models* (see (Earman 1989, Ch2,§ 4)) the laws of physics admit and whether all such models are Laplacian.

To give a taste of what this involves we take the case of a generic dynamical system. It has a state space X. Earman's notion of world-at-a-given-time is simply a point in this space. A world is represented by a time-indexed curve in this space. The deterministic laws of nature are represented as a flow map $\Phi : \mathbb{R} \times X \to \mathbb{R}$ which ensures a unique curve passes through every point—or if you prefer fixes all the nomically possible worlds. Most importantly these curves never fork.

This is almost a realistic description of the classical mechanics of point particles, were it not for the fact classical mechanics is not deterministic! I direct you once again to Earman (*ibid.*, Ch 3) for details why. Lest we think determinism is a thesis without a cause, robust Laplacian determinism holds, for example, in Minkowski spacetime for source-free electromagnetic fields (which describe radiation) (Earman 2007, p. 1395).

What this model does capture is that the laws of classical mechanics for particles are ordinary differential equations. And that flow maps are solutions to those equations (more precisely to the Cauchy problem for these equations). Moving into modern physics we pass from ordinary differential equations to partial differential equations which brings a host of additional technical considerations for determinism. But the basic mathematical idea stays the same: determinism hangs on the existence and uniqueness of solutions (and continuity with respect to starting conditions). These technical issues will not preoccupy us further, except for § 4.5. However as Earman persuasively argues, they will need to be addressed if we are serious about reading the truth of determinism off our most up to date physics.

The moral for philosophers is that determinism is a live scientific concern. And we would do well to ensure our theories are compatible with it, whilst we await for conclusive evidence from physics. Hence the urgency of T1.

I end this section by surveying the way philosophers discuss determinism in the free will debate. Thus let determinism be the claim that

> every true proposition follows, with metaphysical necessity, from the conjunction of any true history-proposition with all the true laws of nature.
>
> (Dorr 2016, p. 242)

with

> If p and q are any propositions that express the state of the world at some instants, then the conjunction of p with the laws of nature entails q.
>
> (van Inwagen 1983, p. 65)

and

> there is a true historical proposition H about the intrinsic state of the world long ago, and there is a true proposition L specifying the laws of nature that govern our world, such that H and L jointly determine what I [do, e.g. raise my hand].
>
> (Lewis 1981, p. 113)

So we all agree the past and the laws jointly conspire to limit present and future

options. There are some differences too, but I will not flesh out whether the differences between metaphysical necessity, entailment and joint determination are substantive.[11] Indeed there seems to be widespread agreement on the premises determinism supports as we shall see in the next chapter. Lewis's definition has the strongest *definiens* as he is explicitly requiring that propositions about the intrinsic (physical) state of the world fix his actions, rather than, say, the positions of all the atoms in his hand etc. As van Inwagen (1983, p. 65) observes "determinism is a thesis about propositions, but the free-will thesis is a thesis about agents" so all the participants in the debate need to bridge this gap somehow with agential abilities to do p providing the key pillar.

2.4 Lewis and van Inwagen

In this section I will briefly go through the standard taxonomy of positions in the free will debate. Here I footnote van Inwagen (2008, p. 330) (and implicitly Lewis).

Firstly, compatibilists deny T1. They defend compatibilism which is the claim that the Free Will Thesis and Det may be true.[12]

Incompatibilists accept T1 and deny compatibilism. They defend incompatibilism which is the claim that the Free Will Thesis and Det cannot.

Soft determinists hold that FWT and Det are true in our world. Thus soft determinists are also compatibilists. Or conversely, compatibilists hold soft determinism may be true, cf. (Lewis 1981, p. 113).

Finally hard determinists hold Det and incompatibilism. Thus they must deny FWT.

Lewis was a compatibilist and van Inwagen is an incompatibilist, but the locus of their disagreement is the latter's Consequence Argument where they both 'feign' determinism. There is a true thicket of related positions philosophers have defended, but, as I am focusing on the debate between van Inwagen and Lewis, this brief taxonomy will be more than sufficient.

Now we are done with motivation and definitions, I move to discussing arguments in the next chapter.

[11] But see (van Inwagen 2004, fn 20) on this point.

[12] Given I have required Det to be a thesis about the actual world, this also restricts the truth of compatibilism to the actual world. I think this is an improvement over van Inwagen's (2008, p. 330) formulation which would allow compatibilism to be false for us but true at some alien world. It also allows me to completely sidestep a recent controversy over whether the Consequence Argument is truly an argument for an 'absolute' version of incompatibilism, for an overview see (Speak 2011, p. 124ff.). Speak's concludes that the Consequence Argument "may not show that determinism necessarily undermines free will, but it does appear to show that the freedom of every human being who has ever existed on our planet would be undermined by it". And this motivates my definition.

Chapter 3

The Consequence Argument

In this chapter I will discuss the Consequence Argument for incompatibilism as put forward by van Inwagen. If sound, this argument goes a long way towards establishing the truth of T1. So much so that van Inwagen (2004, fn. 21) claims it has succeeded in convincing the majority of specialists of the truth of incompatibilism.

We will see it actually consists of a whole family of related arguments. In sections 3.3 and 3.6, I will be focusing on the two which have drawn the most attention. These are what van Inwagen calls the First and Third Formal Arguments.

I will also discuss Lewis' reply to the First Formal Argument (§ 3.4) and his own distinctive brand of compatibilism, namely Local Miracle Compatibilism (LMC). In section 3.5, I give an extended defence of LMC, which is Lewisian in spirit if not in origin (Lewis never spelt it out quite so, but I claim *had he done*, he would have argued along analogous lines).

The purpose of this defence is to uncover the assumptions over which he and van Inwagen disagree without being able to furnish further arguments. This sets the stage for chapter 3 where I will examine Landsman and Dorr's challenges to LMC and assess whether their arguments break the stalemate between van Inwagen and Lewis.

3.1 Timeline of debate

Before engaging with van Inwagen's two arguments I will briefly revisit the timeline of their rise to prominence. The state of play has become quite complex with many replies and counter replies. I cannot hope to to give a full account of the situation here, nor would that be useful. Rather I propose to focus on the original salvoes from van Inwagen and Lewis. Their dialectic is also somewhat tangled so I will briefly revisit it below. Van Inwagen put forward the First Formal Argument in an article (1975). He later included it almost verbatim it in his magnum opus *An Essay on Free Will* (1983), hereafter simply

the *Essay*, where he also introduced his Third Formal Argument.[13]

Lewis issued a rebuttal (1981) of the First Formal Argument as it appeared in van Inwagen's 1975 paper. He did not, to my knowledge, publish any further papers directly addressing the free will question,[14] nor did he reply to van Inwagen's later arguments in the *Essay*. Likewise van Inwagen did not address Lewis rebuttal directly two years later in the *Essay*, but simply gave an updated defence of the First Argument. Finally Lewis's untimely death (O'Grady 2001) prevented him from responding to van Inwagen's direct reply (2004) to his aforementioned 1981 paper.

In summary: we have two arguments put forward by van Inwagen; a strong reply to the first one by Lewis and a yet to be determined dialectical status of the second one. I will have more to say about this later, in particular I will discuss how the premises of these two arguments are related. After all, if van Inwagen has put forward an unanswered argument—let alone an unanswerable one—this should break the stalemate in his favour. I will return to this in §3.6.

I close this section by noting van Inwagen's own view of the situation: we have reached an uneasy impasse as

> there have been no new arguments or ideas of any real consequence. The parties to the discussion of the problem of free will since 1983 know all the relevant arguments and concepts that pertain to every aspect of the problem, and dispute about those arguments and concepts without saying anything that is both new and important about them.
>
> (van Inwagen 2015, p. 25)

Furthermore he has generally accepted the cogency of Lewis's reply (van Inwagen 2008, p. 340, 2015, p. 25), but has not conceded philosophical ground (van Inwagen 2004). He argues maintaining the compatibility of determinism and free will comes at a philosophical cost, like any other worthwhile philosophical thesis. The Consequence Argument helps us see just how *high* it is. Naturally philosophers will differ on whether the "price is worth paying" (*ibid.*, p. 349): Lewis thinks so, van Inwagen does not and perhaps neither do the majority of specialists.[15] So we have a stalemate. My task for the rest of chapter 3 will be to lay bare the claims which bring about this stalemate.

[13] Unsurprisingly, the book also includes a Second Formal Argument which I do not address in the main text. It has not received wide attention. I attribute this to the fact that van Inwagen builds this argument on the notion of an agent 'having access to a possible world'. This relation is *not* the well known accessibility relation of modal semantics, as van Inwagen is at pains to tell us. He cheerfully admits such "'access' talk is artificial" but claims "it not therefore unusable in everyday life." (1983, §3.8). I disagree with him on the latter point, but this is not my main concern. Rather, given my earlier discussion of the gap between commonplace and philosophical notions of free will, I see no reason to widen this gap into a chasm by replacing "ordinary ability talk" with locutions about agential access to possible worlds, as van Inwagen advocates. We may need to resort to modal language to formalize what intuitions we have, but we should not pretend our intuitions come ready-packaged in such form. For these reasons I have set this argument aside.

[14] Although see (Lewis 2015a,b) for his correspondence and published work on the free will defence in philosophy of religion. Of course any such defence *assumes* free will as discussed here.

[15] Cf. (van Inwagen 2004, fn. 21), (Huemer 2000) and (Vihvelin 2015).

3.2 The core problem

The core pattern of the various Consequence arguments can be summed up following Kapitan (2011, p. 2).[16] In the argument schema below and in the remainder of this essay P_0 is a true proposition about the state of the world at some time in the distant past. L is a true proposition asserting the laws of nature and P is a true proposition about the state of the world at t when Alice acts. Thus we have:

1. P is a consequence of (P_0 & L).

2. It is unavoidable that (P_0 & L).

∴ It is unavoidable that P.

The two premises must be supplemented by a crucial inference rule which formalizes the closure of unavoidability under the 'consequence' relation:

UC It is unavoidable that p; q is a consequence of $p \vdash$ It is unavoidable that q.

This pattern of argument draws its strength from two underlying widely shared assumptions. Firstly: nothing can be done to change the past or the laws of nature. Moreover no normal human being has any choice about what they are. Secondly: the present 'follows' from the past given the laws. Moreover this connection is all the stronger under determinism.

The former assumption will come under fire in what follows. We shall see that it figures prominently when assessing the cost of various compatibilist proposals.

The latter assumption will also be scrutinized. The notion of 'consequence' in UC needs to be spelt out in more detail before we can assess its philosophical merit, or the soundness of the argument. We shall see that those who hold a Humean conception of laws (they supervene on the facts, rather than necessitate them) will considerably water down the strength of the connection highlighted by determinism.

Furthermore even if the conclusion is warranted it does not immediately contradict the FWT. However, the claim is that P does specify, amongst *many* other things, what Alice chose to watch. Hence her ability to do otherwise must be defended in the face of the unavoidability of what she actually did. Therein lies the tension which the Consequence Argument brings out.

In order to prove his case the incompatibilist must: i) explain how unavoidability and agential ability are related; ii) explain how the notion of 'consequence' and determinism are related and iii) justify the inference rule UC. I will discuss van Inwagen's attempts to do this in the following sections.

[16] Cf. (van Inwagen 1983, p. 56).

3.3 Van Inwagen's First Formal Argument

This is the First Formal Argument nearly as it appeared in the *Essay*.[17] I have taken two liberties with the original text: firstly I have changed references to the judge J to our Alice (and *her* hand). Secondly, I have changed the phrase 'could have rendered false' into 'was able to render false'. The latter is a substantive change admittedly, but one that van Inwagen would avow, see (2004, fn. 17). It has the dual advantage of preventing confusion: 'could have' can mean 'was able to' and 'might have'; we are interested in the former meaning. It also puts the conclusion of the argument in direct opposition to the FWT which is also couched in terms of abilities. Recall Alice did not actually raise her hand in my scenario. The point of contention is whether she *was able to*: the FWT asserts she was and van Inwagen argues she was not *if determinism is true*.
Without further ado this is the argument:

(1) If determinism is true, then the conjunction of P_0 and L entails P

(2) It is not possible that Alice have raised her hand at *t* and P be true

(3) If (2) is true, then if Alice was able to raise her hand at *t*, Alice was able to render P false

(4) If Alice was able to render P false, and if the conjunction of P_0 and L entails P, then Alice was able to render the conjunction of P_0 and L false

(5) If Alice was able to render the conjunction of P_0 and L false, then Alice was able to render L false

(6) Alice was not able to render L false

∴ If determinism is true, Alice was not able to raise her hand at *t*

It is a valid argument uncontroversially.[18] Hence if we accept the premises, the conclusion follows and, in particular, it follows even though van Inwagen has not provided a worked out theory of agential ability. Hence I agree with Kapitan that "what makes the consequence argument so attractive is that it is plausible *prior* to any analysis of ability" (2011, p. 136), whilst achieving the incompatibilists' main aim.

Therefore a compatibilist must seek to undermine the premises of the argument. I will discuss how Lewis went about this in the next section.

[17] It has a slightly different second premise compared to (van Inwagen 1975) and highlights the distinction between propositions and acts.

[18] Suppose we deny the conclusion: assume determinism and suppose Alice was able to raise her hand at T. Starting with the second antecedent in (3) this sets up a cascade of modus ponens until we infer the consequent of (5), i.e. 'A could have rendered L false' giving a contradiction with (6).

3.4 Lewis's reply: Weak and Strong abilities

I start by summing up Lewis's strategy. His reply consists of pointing out a difficulty with the phrase 'was able to render false'. Specifically he argues that it is a "term of art" (1981, p. 119) and as such needs to be analysed more carefully. He goes on two give two possible meanings to the phrase and to argue that both allow one premise in the First Formal Argument to be denied. He also *hints* that there is no "meaning that would make all [van Inwagen's] premises defensible without circularity" (*ibid.*), but does not give an argument to this end. I highlight this because his reply may win the battle but not the war, as it leaves the door open to further attempts to explicate this crucial phrase. But this part of the debate has displayed all the hallmarks of a war of attrition with neither side gaining a decisive advantage to date.[19] Thus for our purposes it will be enough to restrict ourselves to Lewis's original remarks.
I will now discuss the details of his reply, starting with a definition:

> 'An event falsifies a proposition' iff, necessarily, if the event occurs then the proposition is false.[20]

Agents can also falsify propositions through their actions, which Lewis seems to consider a subclass of events. Thus:

W 'Alice was able to *weakly* render p false' iff there is an action A such that

i) Alice was able to do A

ii) If Alice had done A, then *some event* would have falsified p.

S 'Alice was able to *strongly* render p false' iff there is an action A such that

i) Alice was able to do A

ii) If Alice had done A, then *A or its causal downstream* would have falsified p.

By causal downstream of an event e, I mean any event which has e in its causal history.[21] Lewis compresses both *definiens* into

> 'Alice was able to do something such that, if she did it, the proposition would have been falsified ...'

This is less of a mouthful, but obscures the conditions to be met somewhat. Firstly, I have interpreted clause (ii) in each definition to be a counterfactual claim. This is not clear from Lewis's phrasing. Elsewhere he employs fully counterfactual constructions such as "Had I raised my hand, a law would have been broken" (*ibid.*, p. 116) so I think my reading is licit. Secondly, it obscures the fact that these abilities-to-render-false are parasitic on everyday abilities

[19] See (Speak 2011) and (Kapitan 2011) for the latest news from the front.

[20] Formally: 'an event falsifies p' $:=$ $\Box$('event occurs' $\rightarrow$ 'p is false'), where $\rightarrow$ is the material conditional. I have taken another slight liberty with the original. Lewis has the phrase 'would falsify' in the *definiendum*. I have chosen the 'falsifies' over 'would falsify' to highlight the fact that Lewis is using a counterfactual condition in definitions W and S.

[21] See (1987b) for Lewis's own view of causal histories. For the present purposes it is enough to consider any chain of events in the cause-effect relation with their successor.

through clause (i). It is still a hard philosophical problem to give an analysis of what someone's ability to do A is. As I read him, Lewis thinks he does not need to in order to head off the incompatibilist challenge.

He does so as follows. He insists that our actions may at most be in the causal downstream of a miracle and never in the causal upstream of one. Or, to put it in terms of laws, we are able to render a law of nature false in the weak sense but not in the strong sense. We are able to benefit from miracles but not bring them about. I will assess this claim shortly. First, I will explain how Lewis puts this distinction to work to undermine van Inwagen's argument.

Lewis accepts the first four premises of the First Formal Argument. But he rejects premise (*5*) or (6) depending on which meaning of 'was able to render false' we adopt. And given "there is nothing in van Inwagen's text to suggest any third meaning that might work better than these two" (Lewis 1981, p. 119), the incompatibilist challenge may be met.

On the weak reading of ability, he rejects premise (6). *Some event* would have rendered L false. *Some event* would have broken the law. In his own words:

> Had I raised my hand, a law would have been broken beforehand. The course of events would have diverged from the actual course of events a little while before I raised my hand, and at the point of divergence there would have been a law-breaking event—a divergence miracle, as I have called it (Lewis 1979). But this divergence miracle would not have been caused by my raising my hand.
>
> (Lewis 1981, p. 116–117)

Thus he can maintain, contrary to (6), that Alice was weakly able to render L false.

On the strong reading of ability, he rejects premise (*5*). His reason for doing so seems to boil down to this: van Inwagen has cherry-picked an example that proves his point. Specifically, the exemplar case in the 1975 article is disanalogous to the case at hand. In both cases Lewis maintains the agent is able to render false, in the strong sense, the conjunction of historical and non-historical propositions. In the example the non-historical conjunct is a proposition about someone's future travels. In the First Formal Argument the non-historical conjunct is a proposition about the laws of nature. Hence he disputes what we may infer from this ability in the two cases. Whereas the traveller is able to falsify, in the strong sense, the non-historical conjunct, Alice is not. Therefore the traveller's example does not *support* premise (*5*), it merely provides a case where it is true. The most Lewis is willing to concede is a modified version of (*5*):

*5** If Alice was strongly or weakly able to render the conjunction of P_0 and L false, then Alice was weakly able to render L false.

If he is correct then the First Formal Argument is not sound on the weak reading and not valid on the strong one.

3.5 Arguments for Local Miracle Compatibilism

The plausibility of Lewis's reply turns on whether he has a principled reason to uphold i) miracles may occur (although not in our causal downstream) and ii) the modified premise *5**. I will discuss his reasons in turn in the next two sections.

3.5.1 From counterfactuals to miracles

It is clear Lewis's reply would not be very convincing if his reason for i) amounted to 'it secures the existence of free will'. Moreover, his reply would not be very convincing if this reason were not a *very* good one. He needs to justify the need for a miracle, no less!

Once again it comes down to an issue of philosophical cost. What can Lewis put on the scales to outweigh his belief in miracles? The answer can be found scattered across his writings,[22] starting with his book *Counterfactuals* (2001) and culminating in his counterfactual theory of causation (1973). Counterfactuals certainly bear a large load in Lewis's philosophy. This already offsets part of the cost. It remains to be seen whether miracles are so indispensable to his theory of counterfactuals as to justify his belief in them.
Counterfactuals are subjunctive conditionals with contrary-to-fact antecedents:

N Had Captain Savitsky launched his torpedo, then nuclear war would have broken out.[23]

I do not wish to give a poor rehash of Lewis's theory of counterfactuals, but merely to point out which of its moving parts *requires* miracles. So, in bare outline, a counterfactual '$a \mathrel{\square\!\!\rightarrow} c$' is true just in case the material conditional '$a \rightarrow c$' holds throughout a suitable set of possible worlds centred on the actual one α, with at least one true instance of the antecedent. The suitable set of possible worlds is defined in terms of the relation:

$w\mathrm{S}\alpha$ w is similar to α to degree d or more.

Any world w that makes the cut belongs to the set, any world that misses it does not. Choosing different degrees of similarity leads to a nested class of sets centred on the actual world.

Returning to Captain Savitsky, a small, local change in the course of history would have been enough to bring about an enormous change thereafter. N is a true proposition. However, on a naive, but democratic view—all differences between worlds count the same—N turns out to be false.[24]

[22] See the Appendix of (Lewis 2001) for an annotated list of his writings which deal with counterfactuals.

[23] See (Savranskaya 2005) for a chilling eyewitness account of events aboard the Soviet submarine B59 during the Cuban Missile Crisis. This example is salient precisely because these events *very nearly brought about a nuclear war*. Thus there is no doubt that the corresponding counterfactual is true, in our ordinary practice.

[24] In the actual world the antecedent is false, as is the consequent. Suppose that in w Savitsky

This led Lewis and others (Lewis 2001, p. 75–77, 1979; Bennett 2003, Ch. 12) to refine the similarity relation and give greater weight to certain differences than others. Examples like N show that similarity in the lead up to the antecedent trumps later dissimilarity. And similarity in the lead up trumps dissimilarity in the lead up. So the closer the past matches the better.

This brings out the connection with miracles. Under determinism, perfect match in the past entails a unique future. In turn this entails the contrary-to-fact antecedent cannot be true *given the laws of nature*. If the antecedent were true, a law of nature would have been broken. Thus Lewis must accept this key claim:

M If counterfactuals:

a) require possible world semantics;

b) are compatible with determinism;

c) require perfect similarity in the lead up history;

then a contrary-to-fact antecedent can only be true because of a miracle.

Viewed this way, a miracle is the price we pay for having a theory of counterfactuals under determinism. Counterfactuals pervade Lewis's philosophical system and our own everyday reasoning.[25] Lewis cannot do without them, nor can we. If M is correct then he cannot do without miracles. Thus "a small, localized inconspicuous miracle" (2001, p. 75) is also the *right* price to pay.

To my knowledge he never spelt out the argument in defence of miracles—clause i) above—quite like this, but I think it is the most convincing one he could give to justify his compatibilism. It can also be deployed convincingly against van Inwagen's Consequence Argument because it gives an *independent* reason to doubt one of its premises.

There is a further argument Lewis could add to tip the scale in his favour. Lewis defended a best system account of laws:

> A contingent generalization is a *law of nature* if and only if it appears as a theorem (or axiom) in each of the true deductive systems that achieves a best combination of simplicity and strength.
>
> (*ibid.*, p. 73)

Crucially, this account

> explains why lawhood is a contingent property. A generalization may be true as a law at one world, and true but not as a law at another, because the first world but not the second provides other truths with which it makes a best system.
>
> (*ibid.*, p. 74)

launches the torpedo and nuclear war ensues. Further, suppose that in *w'* Savitsky launches the torpedo, but for some reason nuclear war does not ensue. On the democratic view of similarity, the world *w'* will always be *more* similar to α than any world where the torpedo was launched and nuclear war ensued. The huge change in subsequent history ensures that. Therefore, there cannot be any set around α where the conditional 'Savitsky launched' $\rightarrow$ 'nuclear war ensued' holds. Hence N would be false under a democratic similarity relation.

[25] For an overview of the experimental psychology literature on counterfactual thinking see (Byrne and McEleney 2000; Thompson and Byrne 2002). For a philosophers' view on the uses of counterfactuals see (Bennett 2003, p. 231).

Underlying all of this is a metaphysical thesis, that is "the central empiricist intuition that laws are parasitic on occurrent facts" (Earman 1986, p. 85). Lewis may now haggle thus: The price for miracles is lower than you think. True, laws are never broken here, in the actual world. True, our actual laws are our crutch and we are all the more reliant on them when theorizing in other worlds. But our distrust of miracles need not make the crutch a burden. Lawhood is a contingent matter. Perhaps one small miracle can even be accommodated by laws that are almost our own: that is our actual laws, "complicated and weakened by a clause to permit the one exception" (2001, p. 75). Which amounts to the following. Nearby the actual world, there are worlds where one small, local fact does not fit the pattern of facts we are used to. Nearby the actual world, there are worlds where a small miracle need not entail lawlessness. Rather almost-lawfulness by *our* standards and perfect lawfulness by *theirs*; we can learn something from such worlds nonetheless!

I am sympathetic to Lewis's view on lawhood but this latter argument succeeds only if we also accept a large part of his metaphysics. Firstly, some philosophers (not necessarily incompatibilists) will question his reply precisely *because* they think there are no worlds where the actual laws do not hold. Rightly so, as it stands, Lewis's assertion to the contrary is bald unless you agree with his modal realism. When talking about modalities and counterfactuals we may use possible world semantics. Their usefulness is undoubted when employed in the austere way of philosophical logic. But most philosophers (Bennett 2003, p. 153) have not been willing to endorse the ontology Lewis brings with them.

Often there is still wide scope for philosophical convergence on higher level claims,[26] if we agree to keep a mental asterisk next to all possible world talk. However this is not one of those cases. Here the disagreement over the ontology of possible worlds inevitably spills over into a disagreement about which class of worlds are relevant to our decision making and theorizing.

Lewis believed other worlds were entities very much like our own world: "The other worlds are of a kind with this world of ours" (Lewis 1986, p. 2). If you agree with Lewis that all worlds are equal in some sense, you will have more reason to agree we can learn something from them; to agree that they are relevant to our decision making and theorizing. But if you do not, you will also have reason to doubt the higher level claim. But whilst the disagreement persists, neither reason trumps the other and the impasse persists also.[27] Thus

[26] I note that van Inwagen agrees with me on this point, see his letter to Lewis in (2015b, p. 209).

[27] There is one final haggle to be had over *cost* which I have included out of scruple. When discussing a specific instance of a free action Lewis says (my italics): "To accommodate my hypothetical raising of my hand while holding fixed all that can and should be held fixed, it is necessary to suppose *one* divergence miracle, gratuitous to suppose any further law-breaking" (1981, p. 117).

The situation is the same one as considered for counterfactuals. Take an actual history, allow *one* miracle and then let events unfold lawfully thereon. For counterfactuals that is enough. Each time we want to evaluate one we are committing ourselves to *one* miracle (and no more). However, for free will we are committing ourselves to *many* miracles (one for every free action). If so, Lewis the compatibilist is committed to many more miracles than Lewis the counterfactual logician. Perhaps Lewis would have replied that the miracles do not all take place in the *same* possible world. Thus if we are willing to accept one almost-legal world then we should also be willing to accept a

we come to understand van Inwagen's remark (2004, p. 349), in reply to Lewis, that the Consequence Argument succeeds if it succeeds in raising the price of compatibilism. Especially if it raises it enough that only a modal realist can afford it!

3.5.2 Premise *5* and the principle of disglomeration

I now move to the second prong of Lewis's reply to the Consequence Argument (cf. §3.5). Neither Lewis, nor the authors of a cluster of recent papers on LMC have adequately explained the reasons for denying premise (*5*) in the strong sense and upholding premise *5**, cf. (Beebee 2003; Graham 2008; Oakley 2006; Pendergraft 2011). As I discussed, Lewis flatly denies van Inwagen's example supports premise (*5*). Granted, van Inwagen's example does not work. However, his positive reasons for *5** seem to amount to rhetorical questions:

> Given that one could render false, in the strong sense, a conjunction of historical and nonhistorical propositions (and given that in the cases under consideration, there is no question of rendering the historical conjunct false by means of time travel or the like) what follows ? Does it follow that one could render the nonhistorical conjunct false in the strong sense? That is what would support Premise *5* in the strong sense. Or does it only follow, as I think that one could render the nonhistorical conjunct false in at least the weak sense?
>
> (Lewis 1981, p. 120–121)

The more recent contributions muddy the waters by starting (with the exception of Oakley) from imperfect paraphrases of van Inwagen's original argument. Most of them do not even contain the problematic Premise (*5*) and so cannot properly engage with Lewis's reply. Thus we are left with Lewis's basic contention: we are able to strongly falsify laws and past conjoined without therefore being able to strongly falsify the laws. This might lead you to believe the effort spent haggling in the last section was unnecessary as the goods are spoilt anyway.

Luckily, Horgan diagnosed the problem with (*5*) in (1985). He points out that it follows from two principles about Alice:

D1 If Alice is able to render false the conjunction of two propositions p and q, then *either* Alice is able to render p false *or* Alice is able to render q false.

D2 If p is a true proposition that concerns only states of affairs that obtained before Alice's birth, then Alice cannot render p false.

Thus you can see why I have christened D1 the 'principle of disglomeration'. Admittedly van Inwagen does not explicitly assume D1 and D2 to derive the "general principle" of which (*5*) is a particular case. The general principle is:

multitude of such worlds. And this simply brings us back to the modal realism question. We have come to an impasse once again. I believe the problem lies with miracles *qua* miracles, not in their number.

> If q is a true proposition that concerns only states of affairs that obtained before s's birth, and if s can render the conjunction of q and r false, then s can render r false.
>
> (van Inwagen 1983, p. 72)

Perhaps he believes the general principle stands on its own legs and does not need to be broken down further. However the fact that van Inwagen emphasizes that q is a historical proposition cuts against this reply. Indeed, many would agree that no one can render the past false (as does Lewis in the strong sense). Why focus on q being historical if it does not help his argument? I cannot see how van Inwagen could put this claim to work without relying on D1 implicitly. And thus we have Horgan's diagnosis: D1 relies on "an invalid inference" which in turn undermines (5). I will explain the problem in detail below.

Principle D1 requires the following distribution axiom to hold:

D $\quad \Box(p \vee q) \rightarrow (\Box p \vee \Box q)$.

To see why recall that the antecedent of D1 holds when:

i) Alice is able to do A

and

ii) 'Alice does A' $\boxdot\!\!\rightarrow \Box(\neg p \vee \neg q)$.[28]

The consequent of D1 holds if Alice is able to do A as before and

('Alice does A' $\boxdot\!\!\rightarrow \Box\neg p) \vee$ ('Alice does A' $\boxdot\!\!\rightarrow \Box\neg q)$.

Suppose *arguendo* that Alice is indeed able to do A and that there is a suitable set around the actual world which allows the counterfactual conditionals to be replaced by normal conditionals. By eliminating a superfluous disjunct we have:

'Alice does A' $\rightarrow (\Box\neg p \vee \Box\neg q)$

Having made all these preliminary assumptions D1 can only be true if the following holds:

$$\Big(\text{'Alice does } A\text{'} \rightarrow \Box(\neg p \vee \neg q)\Big) \rightarrow \Big(\text{'Alice does } A\text{'} \rightarrow (\Box\neg p \vee \Box\neg q)\Big)$$

If the counterfactuals were trivially true, that is if Alice does not do A at any world in the sphere, D1 is also trivially true, without further question. If the counterfactuals are to be entertainable, we add the further assumption that 'Alice does A' at some world within the set. This allows us to eliminate the repeated antecedents above. Hence D1 is true if:

$\Box(\neg p \vee \neg q) \rightarrow (\Box\neg p \vee \Box\neg q)$

[28] By applying De Morgan's Law.

is true. In the standard modal logics D is not a theorem (Jennings 1995, p. 9). I take this to be the essence of Horgan's diagnosis.

I think the problem runs deeper. Peter Drábik (2007) has studied what follows from adding D as an axiom, to any normal modal logic.[29] One obtains a logic with extremely strong deterministic properties. In particular D is equivalent (*ibid.*, thm 3.1.6) to the formula:

F $\quad \Diamond p \rightarrow \Box p$.

Now recall $\mathcal{W}$ is the set of worlds which satisfy the laws of the actual world. Let p_t be proposition expressing the physical state of the Universe at t in α. Given p_t is possible—it is true in the actual world—then it will be true at all worlds in $\mathcal{W}$. By changing t we could automatically enforce agreement throughout history and ensure determinism reigns. Van Inwagen could try and turn this to his favour by adding the premise

(0) If determinism is true then the distribution axiom D holds.

However I think Lewis can rebut this move, and frankly so should we. Determinism allows for different histories to obtain within a set of worlds which have the same laws. It prohibits worlds which share some segment of history from diverging thereafter. Contrariwise, F outlaws diversity in history a priori. Determinism cannot maintain its empirical respectability under such strictures. Recall Earman's intimation (§ 2.3) that we should decide the truth of determinism by looking at the properties of the 'great' differential equations of physics. These equations simply do not proscribe solutions with initial states that evolve into different states from the actual one. Quite the opposite: the study of uniqueness of solutions is aimed at ensuring unique histories unfold from the plenitude of allowed starting points!

This concludes my discussion of the arguments for and against premise (5) taken in the strong sense. It also concludes my discussion of Lewis's contribution to the free will debate. In the next section I will discuss van Inwagen's Third Formal Argument which Lewis never replied to. I will show how it is related to the first one and suffers from related weaknesses.

[29] Normal modal logics are a wide class of modal logics and include most of the systems which are generally discussed in the philosophical literature. So they would be the first logics we would look to anyway if we wanted to salvage D1. For a complete overview see (Hughes and Cresswell 1998, pp. 359–362) where the various normal logics are built up axiomatically from the weakest modal logic **K**.

3.6 Van Inwagen's Third Formal Argument

This is van Inwagen's Third Formal Argument:

(1) $\Box\big((P_0 \,\&\, L) \to P\big)$

(2) $\Box\big(P_0 \to (L \to P)\big)$

(3) $N\big(P_0 \to (L \to P)\big)$

(4) NP_0

(5) $N(L \to P)$

(6) NL

$\therefore$ NP

Van Inwagen introduces a new propositional operator N which translates as:

NP "no one has, or ever had, any choice about whether" P holds.

The argument is valid once we grant the following two "inference rules" (van Inwagen 1983, p. 94):[30]

(α) $\Box P \vdash NP$,

and

(β) $N(P \to Q), NP \vdash NQ$.

The following rough and ready 'derivation' should give an idea of the relation between the First and Third Formal Arguments.[31] It suggests they are in fact equivalent.

Start with the disglomeration principle D1 of the previous section as an axiom. Let RP stand for the ability to render P false and D1 becomes

D1 $R(P_0 \wedge L) \to RP_0 \vee RL$.

Contraposing we get:

$\neg RP_0 \wedge \neg RL \to \neg R(P_0 \wedge L)$.

And assuming:

$\neg R \equiv N$,

we obtain the 'principle of agglomeration' as an inference rule:

A1 $NP_0, NL \vdash N(P_0 \wedge L)$.

[30] One wonders whether premises by any other name would appear as true...

[31] See the Appendix of (Horgan 1985) for a more careful study of the relation between the two arguments and in particular between N and R. My derivation captures the key ideas of this discussion.

The principle of agglomeration for N has been widely recognized to be a fatal problem for rule (β), starting with (McKay and Johnson 1996) and ending with van Inwagen (2005, p. 165) conceding the invalidity of (β). I note that this is based on independent counterexamples developed by those authors; the fact that D1 and A1 seem to be different aspects of the same principle is a further problem in light of my discussion in the previous section.

So much so that this has sparked a "cottage industry" (Speak 2011, p. 118) of counterexamples and attempts to amend (β) to get around them. I do not share these authors' faith that more and more fancy versions of the inference rule will unlock hereto untapped philosophical riches, so I will not pursue this line further. For me, it is enough to have shown that the First and Third Formal Arguments belong to the same family and suffer from related difficulties as a result.

3.7 Moral of Chapter 3

Van Inwagen's Consequence Argument is a formidable tool in the incompatibilist repertoire. It brings out the most controversial part of what compatibilists believe and most importantly seems to shift the burden of proof into their court as they must now find and defend an analysis of ability which can escape its strictures.

In this chapter I have focused on Lewis's reply in the form of Local Miracles Compatibilism. This thesis has much to recommend it: it grants most of the incompatibilists' premises, including the definition of free will. It also does not become embroiled in the debates over T3 I alluded to in §2.1. Perhaps it even vindicates the folk view of decision making discussed by Knobe.

The one major downside is its reliance on miracles. I have shown how this assumption is entangled with a host of other considerations across the length and breadth of Lewis' metaphysics. This is cuts both ways for Lewis, it justifies *his* position, but leaves him without allies in defending it.

Finally we now see what brings about the stalemate between him and van Inwagen as we set out to do. In the next chapter I re-examine this balance in the light of new evidence.

Chapter 4

New Challenges to LMC

In this chapter I will discuss two new challenges to Lewis's Local Miracle Compatibilism put forward by Landsman (2016) and Dorr (2016). What distinguishes these challenges is that they both develop novel arguments on physical grounds, the former from quantum mechanics and relativity, the latter from statistical mechanics. This sets them apart from the philosophical literature on LMC we have encountered in the previous chapters. These two challenges have both appeared within the last year and to my knowledge neither have been replied to in print.

4.1 The Free Will Theorem

Once again I begin with some history. The Free Will Theorem is a result originally published by Conway and Kochen (2006) and then further strengthened in (2009). I will be focusing on the reformulation of this result by Cator and Landsman (2014). I will briefly explain why below.

Despite its name, the original Free Will Theorem is first and foremost a result in the foundations of quantum mechanics, in the wake of the well known theorems of Bell and Kochen-Specker. Many have thought that the significance of these results goes beyond the technical debates in quantum mechanics and tells us something about reality itself, in this instance about free will. We live in the age of experimental metaphysics, as Shimony put it (1993, p. 64).

Some of the more sweeping pronouncements of this age will raise philosophers' hackles; certainly Lewis's:

> maybe the lesson of Bell's theorem is exactly that there are physical entities which are unlocalized, and which might therefore make a difference between worlds—worlds in the inner sphere—that match perfectly in their arrangements of local qualities. Maybe so. I'm ready to believe it. But I am not ready to take lessons in ontology from quantum physics as it now is. First I must see how it looks when it is purified of instrumentalist frivolity, and dares to say something not just about pointer readings but about the constitution of the world; and when it is purified of doublethinking deviant logic; and—most of
>
> *cont.*

> all—when it is purified of supernatural tales about the power of the observant mind to make things jump.
>
> (Lewis 1987c, p. xi)

The problem is that there is no agreed upon interpretation of quantum mechanics. And that is precisely what would be needed to settle broader metaphysical questions of interest to philosophers. Thus to make use of quantum mechanics to draw sweeping metaphysical conclusions begs the question.

I agree with Lewis insofar as we should not mistake the metaphysics of any particular luminary—Bohr or Einstein or the author of your favourite textbook—for the metaphysics of the *theory*. Quantum mechanics simply does not come with a metaphysical instruction booklet attached. Nonetheless I believe these results can be a starting point for fruitful philosophical work. As Wüthrich notes, this kind of foundational theorem, properly understood,

> does not in any way rely on an interpretation of the theory, i.e. it does not presuppose a solution to the measurement problem. Just as Bell's theorem and the Kochen-Specker theorem it can thus be seen as imposing constraints on any viable interpretation.
>
> (Wüthrich 2011)

With a viable interpretation in hand and any metaphysical assumptions out in the open, physics has plenty to contribute to metaphysical questions. Perhaps not settle them, but certainly constrain the workable interpretations by suggesting hitherto unforeseen possibilities or difficulties. Rightly interpreted the Free Will Theorem does just this.

This brings me to the Cator and Landsman paper which adopts this attitude from its the very title. They have also zeroed in on the main deficiency of the Conway and Kochen papers. Conway and Kochen's philosophical conclusions are simply unwarranted given their metaphysical assumptions. How so? One of their premises boils down to

EFW Alice has 'free will' in a special scenario:[32] she has a free choice amongst alternative courses of action.

They offer a rather thin analysis of 'free will' which has attracted much criticism.[33] We are told that Alice's choice of action being free

> means [...] that it is not determined by (i.e., is not a function of) what has happened at earlier times (in any inertial frame)
>
> (Conway and Kochen 2009, p. 228)

It should be clear that this is not the Free Will Thesis of previous chapters. Nor is determinism equivalent to functional dependence of state on time as Russell pointed out over a hundred years ago (1912).[34] And above all, if Alice's choice

[32] I have stripped the technicalities to make the philosophical point clear. Nonetheless you may want to check I have not done so tendentiously. The premise I am referring to is MIN (Conway and Kochen 2009, p. 228). The alternative courses of action are choices of mutually exclusive experimental settings—triplets of directions—in a bipartite quantum system.

[33] See (Landsman 2016, pp. 2ff) for a survey of critical literature.

[34] See also (Earman 1986, Ch.2, § 5).

at t can be paired with more than one earlier state of the Universe, then EFW collapses into assuming indeterminism at the outset.[35]

Nothing Conway and Kochen have to say convinces me otherwise. Hence I set them aside in favour of the Cator and Landsman result as it is explicitly aimed at bridging this deficiency. They eliminate the problematic assumptions of Conway and Kochen's version and prove the theorem under the assumption of determinism.[36]

Finally I stress this for future reference: Conway and Kochen's Free Will Theorem and Cator and Landsman's Free Will Theorem go by the same name, but engender almost diametrically opposed philosophical conclusions. I have explained why I reject the former; here on after I will be dealing exclusively with the latter. Hence I will refer mainly to the follow-up paper by Landsman (2016) because it engages the 'traditional' philosophical literature on free will. In particular Landsman argues the Free Will Theorem tolls for Lewis's Local Miracle Compatibilism. If so, nothing could be better, physics finally brings clarity to a long standing philosophical problem! In what is to follow I will play devil's advocate in order to assess the true import of this revised Free Will Theorem.

4.2 Landsman on the Free Will Theorem

The Free Will Theorem comprises four premises and a background 'experimental' set-up. The set-up describes an experiment which has not yet been performed in practice, but may well be one day. It runs as follows.

We consider a bipartite system: two spin-1 particles are produced together in an entangled state and then go their separate way towards distant labs. Alice makes three consecutive measurements on her particle in her lab, as does Bob on his particle in his lab. Both Alice and Bob have several alternative settings available to them, each setting specifies how the three measurements will take place.[37] Both labs use the same type of apparatus and are measuring the same physical quantities on their respective particles. The apparatus returns a reading at the end of the measurement process in the form of a triplet of zeroes and ones, e.g. $(1, 1, 0)$.

Let the set of possible physical states of the Universe be X. Likewise, let the set of possible settings be X_A for Alice and X_B for Bob. Following the original, I will speak of 'settings' laxly using two meanings interchangeably. In the usual physics sense they are just labels for the 'ready to measure on setting s' states of

[35] Cf. (Wüthrich 2011, p. 387).

[36] I will discuss to what extent their *Determinism* premise is indeed full-blooded determinism. I also note that the Free Will Theorem takes the form of a reductio of the Free Will Thesis, thus a claim determinism holds can take appear as an explicit premise (we could very well put it as an antecedent to another premise and derive a contradiction conditional upon it by a valid argument).

[37] It emerges from the proof it is enough to imagine that Alice and Bob each pick a setting from a set of 33 mutually exclusive possible settings: each one is a triplet of directions along which one of the three sub-measurements will take place. See (Cator and Landsman 2014, p. 789) and (Wüthrich 2011, p. 381) on this point.

the apparatus. In the extended sense they are the short chain of events which bring the apparatus into the appropriate label state following Alice or Bob's choice some time before the actual measurement.[38] The content of the set X_Z is trickier to specify, but its nature will emerge in due course. For now, it includes the quantum mechanical state of the bipartite system Ψ_0 and any other "relevant physical variables excluding Alice or Bob" (Landsman 2016, p. 10). Finally, let the set of possible readings, which by design is the same for both Alice and Bob, be denoted Λ.
Now we consider generic functions defined as follows:

$$A\colon X \to X_A, \quad B\colon X \to X_B, \quad F\colon X \to \Lambda, \quad G\colon X \to \Lambda, \quad Z\colon X \to X_Z.$$

We also consider two further functions

$$\widehat{F}\colon X_A \times X_B \times X_Z \to \Lambda, \quad \widehat{G}\colon X_A \times X_B \times X_Z \to \Lambda$$

I may now state the four premises of the theorem:

Determinism The following hold:

D1 The physical situation in the experiment fixes the exact form of the functions $A, B, F, G, Z, \widehat{F}, \widehat{G}$.

D2 For each state $x \in X$ the functions above are *consistent*:

$$F(x) = \widehat{F}(A(x), B(x), Z(x)) \tag{4.1}$$

$$G(x) = \widehat{G}(A(x), B(x), Z(x)) \tag{4.2}$$

Freedom For each $(a, b, z) \in X_A \times X_B \times X_Z$ there exists an $x \in X$ such that $A(x) = a, B(x) = b, Z(x) = z$.

Nature The following hold:

N1 The set Λ contains three possible results is given by

$$\Lambda = \{(1,1,0), (1,0,1), (0,1,1)\} \tag{4.3}$$

(due to the starting *quantum mechanical* state being $\Psi_0 := \big(|-1\rangle\,|-1\rangle + |0\rangle\,|0\rangle + |+1\rangle\,|+1\rangle\big)/\sqrt{3}$ and the quantities being measured).

N2 Bob chooses the same settings as Alice.

Locality In the experiment the function $\widehat{F}$ does not depend on b and the function $\widehat{G}$ does not depend on a (due to the distance between Alice and Bob's lab).

The Free Will Theorem is the result that the above four premises are jointly inconsistent (Cator and Landsman 2014, Thm 4.2). The philosophical conclusion is that this result "challenges compatibilist free will á la Lewis" as it pits

[38] Ultimately we are interested in the moment Alice chooses a particular setting; presumably this event occurs somewhat earlier than her apparatus being set to a label-like value we call 'the setting'. But this is akin to Alice raising her hand to choose a movie. We speak of the action choosing a movie, of her raising her hand to bring this about *ceteribus paribus* and presumably this is all preceded by some event in her brain we might ultimately wish to call 'her choice'.

two of its central claims against modern physics in the form of the latter two premises above (2016, p. 1, p. 11).

The premises above need some further unpacking before they can be explicitly related to Lewis's views. As they stand both *Freedom* and *Determinism* are not just alternative formulations of FWT and Det from Chapter 2. Thus substantive question is whether i) Lewis committed to these two premises and ii) whether he can circumvent this argument in some other way. I will examine both these points in the following sections.

4.2.1 Interpreting the *Determinism* premise

The first issue *Determinism* brings up is this: what *are* the functions posited by D1 exactly? In particular how does the physical situation *fix* their exact form?

Firstly, Landsman's own words on the matter (when referring to the function Z): it "should be chosen (by the theory in question)" so that D2 holds. I have interpreted this more widely in D1 by explicitly requiring that the physical situation fixes their exact form. Thus both the laws and the particulars of the situation contribute to fixing these functions. This is not simply nit-picking. Given two sets, a multitude of functions exist between them in the austere mathematical sense. We need a criterion to pick out the relevant one, and, more to the point, the one that we may put to philosophical work. As I read him, Landsman is thinking along the same lines when he stipulates in a footnote that the physical state x should "control what is going on, such as Alice's actions, namely through functions like A. Without these states[39] mean little" (*ibid.*, fn. 6).

We will see none of these functions are fixed in a straightforward way by current physical theory, but for a different reason in the case of $\widehat{F}$ and $\widehat{G}$ as opposed to A, B, F, G, Z. Nonetheless it is enough that they could be in principle for the argument to succeed.

The interpretation of QM and the functions $\widehat{F}$ and $\widehat{G}$

In this bipartite experiment we start with the entangled particles in a particular *quantum mechanical* state Ψ_0. This is the complete state of the system, at least as far as for-all-practical-purposes (FAPP) quantum mechanics is concerned. Where I say FAPP, many would say orthodox or standard, but I specifically do not wish to prejudge issues of interpretation. Given this state, FAPP quantum mechanics does not fix the outcome as a function of settings and starting state. Indeed the notorious Collapse postulate is required to ensure that the theory places the state of the system *after* measurement in the correct set $\mathcal{O}$ of 'outcome states'. Even then the theory does *not* fix which outcome state the system is in, except in the special case where $\mathcal{O}$ contains just one possible outcome. When multiple outcomes are compossible with the quantities being measured

[39] For absolute completeness I note that Landsman is referring to Alice's *psychological state a*, which in the bipartite experiment presumably bring about the *setting a* (see the discussion leading to footnote 38).

the best FAPP quantum mechanics can do is predict the *mathematical* probabilities of each outcome occurring (the Born rule).[40] Yet further interpretation is required to link them to the relative frequencies of outcomes in repeated experiments, which is what we can observe.

For our state Ψ_0, the set of outcomes $\mathcal{O}$ contains three states because of the quantities being measured. In FAPP QM this also means that Λ contains three outcomes as stipulated by *Nature* because the apparatus is assumed to correlate perfectly with the system it is measuring. Thus the functions $\widehat{F}$ and $\widehat{G}$ are not fixed by FAPP quantum mechanics. However it would be wrong to surmise that the measurement process is the locus of indeterminism in quantum mechanics under *all* viable interpretations.

Certainly, one often hears determinism fails because of what we know from quantum mechanics. So much so that this idea has permeated through to philosophers engaged in the traditional free will debate, we even find a salient example in the *Essay* (1983, p. 191ff). However this failure lies with the philosophy of physics PR department rather than at determinism's door. The question of determinism in QM is inherently caught up in the problem of interpreting the theory as I have been at pains to point out. In particular the thorniest issue centres on the problem of interpreting what the theory has to say concerning measurements. Zooming in even further, one of the central points of disagreement centres around the interpretation of probability in the Born rule: is it objective or epistemic? Does it refer to individual systems or to ensembles of identical systems or even to actuarial classes of relevantly similar systems? Is it an intrinsic single-case property of the system?

A full discussion of these issues would lead us too far into the depths of the debate about the interpretation of quantum mechanics. I can safely say no consensus has yet been reached. Nonetheless I recommend appendix A of (Leifer 2014) for a crisp overview of the main positions on the question of probabilities. My own position is I agree with Leifer that objective intrinsic probabilities are incompatible with determinism, as did Lewis (1987a, p. 118). If *interpreted* quantum mechanics turns out to contain such probabilities then it is indeed indeterministic. However I join him in arguing that it is the problem of interpreting the theory, in particular the probabilities in the Born rule, that may lead us to postulate such probabilities and not the other way round. Thus in any debate where the foundations of quantum mechanics are themselves up for grabs, then so is the issue of determinism.

If we move beyond FAPP QM, to QM* say, the theory may well be able to provide a function that returns the outcome given the settings, given the initial state Ψ_0 and any additional variables it may require (aside from a variable specifying the outcome itself in retro-causal fashion). I hesitate to use the term *hidden variables* for these additional inputs in QM*, but I suspect many would interpret them in this way (and so disdain them also). *Contra* these naysayers, I read Landsman as suggesting that we do *not* try to explicate $\widehat{F}, \widehat{G}$

[40] By mathematical I mean probabilities as axiomatised by Kolmogorov, again without prejudging the question of their interpretation.

in any more detail than we need, whilst rejecting indeterminism during the measurement process. Perhaps, if the additional variables are truly hidden they are inaccessible to us, but they may contribute nonetheless to fixing the outcome through functions like $\widehat{F}$ and $\widehat{G}$.

Supervenience and the functions of physical state

The functions A, B, F, G, Z bring up related concerns and some metaphysical concerns of their own. Take the function A as an example. On the most austere interpretation, it simply associates physical states of the Universe with settings on Alice's apparatus. The physical states of the Universe will be the basic entities in the yet to be discovered Theory of Everything—something like wavefunctions in quantum mechanics or fields in relativistic theories. This may bother you as it is very hard to pin down an entity which is always one new theory away. This is a recognized issue with physicalism going back to Hempel, but I do not think it ought to trouble us unduly.[41] It affects all physicalists equally. Rather, I think the more troublesome issues lie at the level above the fundamental one, where the sets X_A, X_B, Λ, X_Z reside. How are actions, settings and other agential-level states related to the physical state? And what part does this play in fixing the functions above?
For example, Landsman requires that

> A is the function that describes which action $a = A(x) \in X_A$ the agent takes if the world is in state x.
>
> (2016, p. 7)

Further down he adds

> Alice's action a at some fixed time t is determined by the state x of the world at *that time*.
>
> (*ibid.*, my italics)

One way to interpret this is in terms of a supervenience relation, e.g. no change in the setting a without a change in the state x which obtains when the setting is chosen.[42] This approach also commends itself for often being cited as a minimum requirement for physicalists of all stripes (Lewis 1983a, pp. 361–365). On this interpretation D1 says:

D1* The agential level states *supervene* on the physical state of the world: there can be no change in the agential level states without a change in the

[41] Hempel's essay (2001) is usually quoted as the first place this dilemma was posed. This is curious as he was actually discussing biology in the only passage (*ibid.*, p. 192) that vaguely resembles a statement of the dilemma. Further on he discusses a difficulty for determinism which is again related but not *quite* the issue above. All in all, I prefer Crane and Mellor's (1990, p. 188) crisp statement of the problem.

[42] This interpretation is vindicated by Butterfield (2012, p. 9) who mirrors Landsman's second quote when defining supervenience: the higher level property *is determined by* the lowel level one. On the other hand it does exclude the "special case" Landsman mentions in footnote 5 (2016, p. 7) whereby the set of physical states is given by a Cartesian product of the agential level sets.

> physical state. Thus the physical situation fixes A, B, F, G, Z by fixing which a, b, f, g, z supervene on which x.

You could still wonder whether the introduction of the supervenience relation is not simply removing the original question by one step. One can still ask how does the posited supervenience relation fix the functions above? Sadly, philosophical discussions of supervenience are rarely couched in these terms. Two recent exceptions are (List 2016; Yoshimi 2011).[43] In particular, Yoshimi introduces a 'proposition' to the effect that supervenience relations naturally induce corresponding functions.

> If $\mathbf{A}$ and $\mathbf{B}$ are state sets, and $\mathbf{A}$ supervenes on $\mathbf{B}$, then there exists a function $f : \mathbf{B}' \to \mathbf{A}$ (where $\mathbf{B}'$ is a subset of $\mathbf{B}$), defined by the rule $f(X) = Y$ iff X determines Y.
>
> (2011, p. 378)

Nonetheless these general statements of how such functions arise will leave you unsatisfied if you crave details. As far as I can tell, these authors see the existence of the supervenience function as a more formal and correct statement of the basic metaphysical claim supervenience is making. I should also add they place themselves squarely in the the non-reductive physicalism camp. This may explain why they do not feel the need to elaborate on the origin of these functions any further: the levels are primary, whilst the functions are derivative ways of associating them. List suggests we could conceive of genuinely different levels of possibilia—worlds of level-specific facts:

> higher-level worlds need not be identified with equivalence classes of lower-level worlds; they merely pick out such equivalence classes. [...] In fact, two distinct levels Ω' and Ω'' in $\mathcal{L}$ could supervene on one another and could thus be viewed as distinct but isomorphic.
>
> (List 2016, p. 9–10)

Of course taken literally such an ontology would be even more expansive than Lewis's modal realism. The usual alternative is to believe in some form of representationalism about possible worlds (Bennett 2003, p. 155), for example by considering them maximally consistent sets of sentences. In any case, a representationalist will have no trouble in admitting that we use different levels of *description* and that we could allow for this explicitly in our possibilia, without expanding our ontology.

In that case, I would still ask: suppose we are given a supervenience function, can we verify experimentally whether it correctly links different levels of descriptions? Take the example of a table and the phase space of its atoms as the supervenience base for simplicity. The table surely supervenes on the actual arrangement of its atoms. The table 'state' is associated to a point in the phase space and to many, many others besides—it still the same table if I move it a foot to the left or in any other direction. Supposing we had the supervenience function, would we able to check it is indeed correct? In detail: would we

[43] The idea is also foreshadowed in (Sober 1999, p. 137).

be able to check the correctness of each pairing of table to phase space point? *Yes* in principle, difficult in practice. There is a sense in which the function is *up to us*. That is it may well be mind-dependent insofar as what counts as a table depends on social and cultural facts which would in turn depend on what individual people did and do hold to be a table. Physicalism implies (hopes?) that these mind-dependent facts will themselves supervene on what is the case physically. But the cautious physicalist should be prepared to accept that the function she ends up with may be an 'imperfect deserver' of its codomain, or one among such functions.[44]

We have reached deep metaphysical waters, I do not propose to plumb their depth any further for now. Enough that this brief excursion has convinced you not to rule out such supervenience functions a priori. And enough to note that Landsman's argument applies to any supervenience function we may come up with if they are non unique. The substantive assumption, which I join him in making, is that such function *can* exist.

One final question: is the premise *Determinism* a full blooded vindication of determinism? Not quite, nor is this required for the argument. Landsman says clearly it would need to be "supplemented with Laplacian determinism" in order to fully merit its name (2016, p. 10). As I have shown, all we are assuming is determinism throughout the measurement process, albeit in a roundabout way, taking a long detour through the measurement problem in QM and the interpretation of probability. I suppose nervous compatibilists might turn all of this on its head, using the Free Will theorem as a reductio of any interpretation which threatens free will. Or they may reject the metaphysical assumptions I have outlined, starting with the supervenience thesis. I am not amongst them. Nor need they panic just yet: there is still a premise to look at!

4.2.2 Interpreting the *Freedom* premise

My immediate aim in this section is to interpret the *Freedom* premise. But my overarching concern is to find out whether its denial entails FWT is false. Thus I reissue the FWT below adapted to our present situation:

FWT Alice was able, at some time just before the measurement, to raise her hand and thus select a different setting a' to the setting a she actually selected.

We may immediately note that *Freedom* does not mention any abilities. I have argued at length that the Consequence Argument should be understood as an attempt to undermine FWT without committing to any particular analysis of ability. Likewise Lewis's reply is couched in similar terms, he is making claims about cases when can reasonably claim to have a certain ability. He argues we do not have the ability to strongly falsify a law, but compatibilism requires no such ability to succeed. Up to this point I have been happy to join van Inwagen

[44] I borrow this terminology from Lewis (1989, pp. 132ff) who used it in discussing the *far* thornier problem of the supervenience of *value* on psychological facts.

and Lewis in not addressing the issue head on. But now we must confront it, as part of Landsman's philosophical argument addresses the ability question.

Landsman starts by considering a generic experiment (2016, §3). This allows him to introduce his analysis of Alice's ability to do otherwise and relate it to Lewis's views. He uses this framework to argue that *Freedom* does indeed capture one of Lewis's key commitments. Later on, after discussing the Free Will Theorem, he suggests that the bipartite experiment is a special case of the generic experiment and that the latter's philosophical implications hold good (*ibid.*, §5).

So the overall argument has two phases: first establish a point of contact with Lewis in a generic setting and draw its philosophical implications. Second show the Free Will Theorem is a special case of the generic setting. Crudely the conclusion would be: the Free Will does indeed disprove Lewis (or perhaps the feigned soft determinist Lewis (1981, p. 113)). In actual fact, Landsman's argument is much more subtle and nuanced. What is to follow is my own interpretation and critical assessment of the overall argument.

Landsman on abilities and agency

As before, we consider a physical state space X and an agent Alice and her actions $a \in X_A$ in choosing settings for a generic experiment. Landsman assumes that at any given time we may cleanly cleave Alice's *inner* states $i \in X_I$, from what is left over, the *outer* states of the world $o \in X_O$. As before he also assumes the existence of three functions:

$$I : X \to X_I, \quad O : X \to X_O, \quad A : X \to X_A, \quad \widehat{A} : X_O \times X_I \to X_A,$$

subject to the consistency requirement

$$A(x) = \widehat{A}(O(x), I(x)). \tag{4.4}$$

I define *Freedom** below. It is the *Freedom* premise of the Free Will Theorem adapted to the current generic experiment.

*Freedom** $\quad \forall (o, i)\, \exists x : \; o = O(x) \quad \wedge \quad i = I(x).$

I quote the next passage *in extenso*, as it contains a key step in Landsman's argument:

> Lewis wants to make sense of the idea that although (according to determinism) Alice's action $a = \widehat{A}(o, i)$ at some fixed time t is determined by the state x of the world at that time through (4.4) and hence through
>
> $$x = \varphi(x_0, t - t_0) \tag{4.5}$$
>
> it was determined also by any earlier state x_0 of the world at time t_0, nonetheless Alice was able to act otherwise at time t, e.g. she was able to do
>
> $$a' = \widehat{A}(o', i'), \tag{4.6}$$
>
> *cont.*

but did not do so, because doing a' would have illegally modified the state x.

Alice's ability to do a' means that there exists a state x' of the world close to x in that

$$o' = O(x') = O(x) = o, \tag{4.7}$$

making the outer environment in which Alice acts the same as in the actual world, but

$$i' = I(x') \neq I(x) = i, \tag{4.8}$$

where i' should be similar to i in some appropriate sense such that (4.6) holds.

(Landsman 2016, p. 8)

The passage suggests the following analysis of ability, which I have gathered into one *definiens*. Given Alice actually did a, and the states actually were x, i, o. Then

L1 Alice was able to do $a' \neq a$ iff at t there existed a different a'-compatible physical state which determined a different inner state for the same outer state:

$$\exists x' \neq x : \; o = O(x') \quad \wedge \quad i' = I(x') \quad \wedge \quad i' \neq i \quad \wedge \quad a' = \widehat{A}(o', i') \tag{4.9}$$

A difficulty with this interpretation arises in the next paragraph, where Landsman draws the implications of the above analysis for an agent's freedom:

> The point, then, is that according to our *Freedom** assumption, there indeed *is* such a state x', for any given i' and (o, i). Thus the freedom the agent has is precisely what we have formalized as *Freedom**: even *given* the state o of the external influences on her behaviour (and possibly even the state of the rest of the world), there is a different admissible state x' of the world such that, had this state been actual, the agent would have done a'.
>
> (*ibid.*, p. 8)[45]

The difficulty is laid bare if we compare *Freedom** with L1: the former does not mention the action a' and is quantified differently. For now I assume that Landsman agrees with me that freedom and ability to do otherwise are related according to FWT. Therefore, the *definiens* of L cannot be equivalent to *Freedom**, at least *prima facie*. Perhaps, we need to look at the above analysis of ability more carefully in order to determine its logical relation to *Freedom**. I turn to this in the next section.

Logical relation between *Freedom** and L

*Freedom** quantifies over inner states, whilst i' appears as a free variable in L1. This prompts me to depart slightly from the first quotation and revise L1 to consider all possible inner states. I preserve the key idea of the analysis: we wish to find a state that determines the outcome a' whilst keeping the outer state o fixed. Hence, given a, i, o, x at t in the actual world.

[45] I the original passage Landsman refers to *Freedom* meaning my *Freedom**. I keep the labels distinct because I will later be discussing whether *Freedom** is indeed a special case of the *Freedom* premise of the Free Will Theorem on page 30.

L Alice was able to do $a' \neq a$ at t iff

$$\exists(x' \neq x, i' \neq i): \quad A(x') = a' \quad \wedge \quad O(x') = o \quad \wedge \quad I(x') = i'. \tag{4.10}$$

I argue L is the best interpretation of Landsman's analysis of ability. Now to the question of its logical relation to *Freedom**.
Suppose *Freedom** holds. Let $X_{a'}$ be the set of a'-compatible physical states:

$$X_{a'} := \{x' \in X : A(x') = a'\}. \tag{4.11}$$

Note the actual $x \notin X_{a'}$ by construction. Let X_* be the set of (o, i')-compatible states induced by *Freedom** for fixed o:

$$X_* := \{x \in X : \ O(x) = o, I(x) = i' \text{ for all } i' \in I \setminus \{i\}\} \tag{4.12}$$

Note a given $x' \in X_*$ does not necessarily determine the a' we want.

We may now compare the a'-compatible physical states with the (o, i')-compatible ones. On this reading, *Freedom** implies X_* is non-empty. Likewise, the *definiens* of L is the claim the intersection

$$X_{a'} \cap X_* \tag{4.13}$$

is non-empty. On this reading, X_* being non-empty is certainly a necessary condition for L to be satisfied. But it is *not* the case that *Freedom** is a necessary condition for L to be satisfied. Indeed its denial entails some particular $(o^\dagger, i^\dagger)$ pair has no supervenience base for all x. But that does *not* mean the intersection in equation 4.13 is empty, indeed $o^\dagger$ need not even be the actual outer state o. Finally I am open to the possibility that *Freedom** is a sufficient condition for L being satisfied. As I discussed earlier, we do not have access to the physical state x. We do have access to our own inner states and observations about the world on the agential level. So if I claim I was able to type this sentence in Italian, it is because I know I could have been thinking in Italian, stringing together words in Italian, in short I could conceive of the right inner states without changing the world outside. And thus I could also convince myself the right pattern of neuronal activity existed (and could have been actual) and thus the right physical state existed too.[46] In short, if we can figure out plausible agential-level states the physical state can usually be found too. So I am open to the possibility that L is satisfied *whenever* X_* is non-empty.

But all of this does not help Landsman's original argument as I have interpreted it. Zooming out for a moment, recall this: the upshot of the Free Will Theorem is that we should deny *Freedom**. Thus, to undermine Lewis, we must show that *Freedom** is at least a necessary condition for one of his views. Thus even showing it is a sufficient condition does not help.

Perhaps I have read Landsman wrong, or perhaps he thinks "the freedom the agent has" in the second quotation is not to be equated with the ability to do otherwise and other considerations come into play. But we can agree that

[46] I will have more to say on the connection between what we can conceive of or imagine and what is possible shortly.

there must be an appropriate link between Lewis's views and *Freedom** for the consequences of the Free Will Theorem to bite.

I maintain Lewis equated freedom and ability to do otherwise as laid down by FWT and *that* is the locus of the debate with van Inwagen. Does this mean I ultimately believe the conclusions of the Free Will Theorem are not relevant to Lewis's compatibilism? No. Does it mean I reject the analysis of ability in L (or L1)? No. I simply think the fault lies elsewhere and that this generic framework has not revealed it. In the next section I will return to the Free Will Theorem and the bipartite experiment to explain why I think it creates a difficulty for Lewis .

4.2.3 *Freedom* interpreted

In this section I seek to identify a claim Lewis is committed to which is undermined by denying Freedom. I take this as the fundamental constraint on interpreting the *Freedom* premise and, to the extent we share this end goal, the justification for departing from Landsman's original argument.

Suppose we deny Freedom then: there exists an *(a,b,z)* triplet such that no physical state x determines it. Or more precisely: no x determines Alice's setting a, Bob's setting b and the external state variable z jointly. Going back to the supervenience picture, this sounds highly suspicious. We have found settings with no counterpart in physical reality, physicalism is wrong after all!

If we insist certain agential-level states are possible, then we pave the way for this conclusion. I am not ready to accept it, at least not without re-examining the assumptions that lead to it first. This brings us back to the definition of X_A, X_B and X_Z. How did we determine their elements? As denizens of the actual world all we have access to are actual world settings and actual world observations. In our scenario, we may have used all the available settings during various different actual experiments, but does that imply these settings were all genuinely available during each run? If Alice's apparatus has a knob with ten marked settings, are there ten *possible* $a \in X_A$? Or are there ten *allegedly* possible marked settings? If the latter holds denying *Freedom* holds no paradox for supervenience physicalism. If the latter holds and *Freedom* is false, we were simply wrong about what is possible. We were wrong about what settings were genuinely available during a given run.

The principle of recombination of possibilities

By arguing for this distinction between what is possible and what is allegedly possible, I am highlighting a step that we normally do not think twice about, except here it matters. As Lewis (1986, p. 113) put it "how *do* we know" possibilia, when we have no causal (or physical) access to them?

Lewis's reply to this question has two parts: a foundational one and a practical one. I forgo discussing the first part as it does not concern me here. When Lewis answered it, his primary aim was to rebut those who use the question

as a reductio of his modal realism—roughly, how can you know something (possible worlds) that by construction you cannot access?
I quote directly from the second part of his answer:

> Or I can take the question of how we know as a request for 'naturalistic epistemology'. Never mind what makes our modal opinions count as knowledge; how do we come by the modal opinions that we do in fact hold? ('You say the dollar will be devalued tomorrow—*how do you know*?' Imagine the question is asked *not* by a doubter or an epistemologist, but by an official seeking your help in plugging leaks. He wants to know how you came to think so.) [...]
> I think our everyday modal opinions are, in large measure, consequences of a principle of recombination.
>
> (Lewis 1986, p. 113–114)

The principle of recombination is the requirement that "patching together parts of different possible worlds yields another possible world" (*ibid.*, p. 87). This is another metaphysically 'deep' thesis whose full import I do not wish to sound here. Enough that Lewis holds it. I do wish to put it work in our bipartite experiment. So we may match together parts of different possible experiments—say different *actual* experiments—to yield another possible world. We may match together different actual settings chosen by Alice and Bob, with different external states, to yield another possible world. Or put differently, given a combination of settings and external states an appropriate world exists. For each a, b, z there exists a complete physical state of the world x such that x determines that triple. There we have it: *Freedom* is a precisification of the principle of recombination above.

Lewis himself argued along these lines when discussing mathematical *representations* of possible worlds as n-tuples of real coordinates a lá Quine (1968, p. 10ff):

> For every Quinean ersatz world, there is a genuine world with a represented pattern of occupancy and vacancy. This is just an appeal to recombination. But we are no longer applying it to smallish numbers of middle-sized things, horses or horns of heads. Instead we are applying it to point-sized things, spacetime points themselves or perhaps point-sized bits of matter or of fields.
> *Starting with point-sized things that are uncontroversially possible, perhaps because actual, we patch together duplicates of them in great number (continuum many, or more) to make an entire world.*
>
> (Lewis 1986, p. 90–91, my italics)

I cannot claim such a triumph for physics in good conscience. It would undermine modal realism *tout court* let alone free will. We deny *Freedom* on physical grounds—it leads to contradiction given minimal quantum mechanics and relativity theory. On the other hand, the only limit Lewis imposed on possibilia is logical consistency. There are no worlds where p and not-p hold, but otherwise pretty much "anything can coexist with anything else".[47] It may be a

[47] Well not *quite* according to Lewis, but the problems he has with this paraphrase are distinctly philosophical concerns, nothing hangs on them here, cf. (1986, p. 88ff).

nomological contradiction to suppose *Freedom* but it is not a logical one. And so Lewis could argue our bipartite experiment does not supply the correct grounds for giving up the principle of recombination in the form of *Freedom*. The laws could be different—he believed they are parasitic on different possible facts—and thus the principle of recombination is not subject to them. I agree with him insofar as I think the Free Will Theorem does not settle the truth of modal realism. However Local Miracle compatibilism is not in the clear yet.

Lewis, Landsman, van Inwagen and anyone interested in the truth of compatibilism share two implicit assumptions. Firstly that the actual world is a physical object, the basic properties or entities are the ones we learn (or will learn) from physics. Were it not so, we would hardly be troubled by determinism.[48] Secondly that the relevant alternative worlds when discussing agent's actions are also physical objects (or representations of arrangements of physical objects).

Based on these shared assumptions we could adopt the following weak constraint on ability:

N If Alice is able to do action a then it is nomically possible to do a.

Thus nomic opportunity to do otherwise—the actual laws do not proscribe it—is a necessary condition for ability to do otherwise. When the action is specified in terms of entities mentioned by the physical laws N can be checked easily.[49] Conversely, when the action is specified in terms of agential-level states and objects it may be harder to check this criterion. It is also easier to be mistaken about what the laws allow. This brings us back to Alice and her choice of settings.

When we are considering what Alice is able to do we must restrict ourselves to worlds relevantly like our own. Hence we must have some way of delineating the relevant alternatives. We could adopt the following agential-level principle of recombination to take this into account:

PR1 Her options are parts of (sets of) worlds where any setting a can coexist with any b and z as long as the actual laws do not say otherwise.

Alice cannot consider a putative choice of setting an option if it then turns out this setting would violate the laws if realized. Given compatibilists *do* want to say she is able to select any of the marked settings on the apparatus we

[48] Agreed it is more complicated than this and I would do well to heed 'Hamlet': *there are more things in philosophy than are dreamt of by me...* In that case treat it as a constraint, I am only interested in addressing this type of compatibilists.

[49] Suppose I claim I am able to strip an atom of its outer shell electrons. Then according to N there exists a configuration of the atom and the electrons which would bring about that possibility. We can easily see this is indeed nomically possible as laws rarely proscribe any given state if we allow any history. I am *not* saying: It is not possible for me to be in Proxima Centauri in ten minutes given my history up until now locates me on Earth. Had I been elsewhere it might have been nomologically possible. I am saying: it is not possible for me to travel at twice the speed of light.

must also suppose all the marked settings are nomically possible. Thus *Freedom* formalizes a necessary condition for Alice to have the ability to choose any marked setting. In short: Freedom is a necessary condition for a claim compatibilists want to make.

No doubt van Inwagen would happily endorse this principle. Lewis, on the other hand, could disagree and argue:

PR2 Her options are parts of (sets of) worlds where any setting a can coexist with any setting b and z as long as the actual laws or almost-the-actual laws do not say otherwise.

On Lewis's version, *Freedom* is not a necessary condition for his compatibilism unless we can give an independent reason to exclude the almost lawful worlds from being relevant. You could read him as arguing for a slight weakening of condition N above.

Thus the debate between van Inwagen and Lewis can be recast as a disagreement about which subclass of these physical worlds is relevant: van Inwagen will ultimately insist on those with exactly the same laws as ours, Lewis will allow these *and* further physical worlds which are almost exactly lawful (by the our standards). To this extent you might conclude there is nothing new under the sun, this disagreement has already been revealed in the free will literature on ability:

> One way to see the disagreement between incompatibilists and compatibilists about determinism and being able to do otherwise is as a disagreement about what worlds are relevant. According to incompatibilists, all and only worlds with the same past and natural laws as W are relevant; they hold the past and the laws fixed. Compatibilists disagree.
>
> (Mele 2003, p. 451)

It would be a shame to have come this far and be left with this quietist conclusion. So note this: *contra* Mele, the Free Will Theorem takes *most* of the past out of the reckoning.[50] Now the disagreement centres on laws versus almost-laws, i.e. PR1 versus PR2. This is already an advance.

Going further, can we adjudicate between the two principles? If one small miracle at just the right time allows Lewis to sidestep the Consequence Argument, can he do the same with the Free Will Theorem?

This depends on whether the Free Will Theorem creates further difficulties in accepting almost-lawful worlds as relevant. Recall: in the almost exactly lawful worlds Lewis argued an agent's actions could *not* cause a miracle, merely follow from one. That was the basis of the distinction between weak and strong abilities. This means there is a constraint on the almost lawful worlds I can try to exploit.

Does Alice cause a miracle if her choice corresponds to an unlawful triplet (a, b, z)? This is a delicate question. Until now I have been silent on the issue

[50] I say most of the past because the *Determinism* premise and the setup of the experiment *at most* bring into consideration events very shortly before the measurement. In this sense, not assuming full blown Laplacian determinism is a strength of the argument.

of how exactly the perfect correlation of Alice and Bob's choices in *Nature* is enforced. One way is to suppose Alice chooses settings for both wings ahead of time and perhaps asks her colleague Bob to enact them (assuming he does so without fail). In that case, the occurrence at almost lawful worlds of an unlawful (a, b, z) triplet follows in time from Alice's earlier actions. And you could also insist she caused it to undermine Lewis. Alternatively, Bob does choose, in the same determined fashion as Alice, and they both do so when they are too far apart for a signal to pass between them. What then? Can we say one side acted first in an absolute way? If not, as special relativity teaches, then Lewis could exploit the gap and say Alice did not cause an unlawful triplet to come about. Notice in both cases I have said nothing about what is required for causation in the hope all parties would agree on specific instances if not on the general analysis. Equally they might not, with the disagreement in this scenario reflecting a disagreement in the general analysis. I will not try and settle these questions. I suspect there is room for improvement on my analysis of the bipartite scenario from the philosophy of physics quarter. This might help clarify some of the issues around causation.

Nonetheless, Landsman and I are implicitly committed to siding with van Inwagen, if the Strong Free Will Theorem is our only tool. Its conclusions are restricted to states x where the laws of physics are as they actually are. And even if Alice causes miracles by her choices in the bipartite experiment, Lewis can still retreat without conceding defeat across the board. Firstly, denying Freedom entails at least *one* a, b, z triplet is problematic, or in my earlier phrasing, only allegedly possible. Suppose *arguendo* we keep z fixed, then Alice and Bob can each choose from a set of 33 settings (the proof uses that many, perhaps the minimum number is smaller). Lewis could accept one particular alternative setting a' is not allowed and maintain Alice was able to nonetheless select otherwise, she was able to select a'' if not a' from the remaining settings.

I would reply that not being able to select one alternative a' ends up being as bad as not being able to select any alternative settings on symmetry grounds. Very roughly put: the laws of physics have the same form in all directions; the experimental settings boil down to triplets of directions along which a measurement is made; therefore there is no reason to suppose one direction is special. I would not be surprised if many physicists are sympathetic to this line of argument, as I suspect it underlies the intuition that all settings are available in some sense. Once again philosophy of physics could come to the rescue, by clarifying whether one or many settings are proscribed and whether there is any merit to the latter 'symmetry' argument. My own hunch on the latter is it does not in its current form. The premise *Nature* requires a specific initial state and that breaks the symmetry.

4.3 Moral of the Free Will Theorem

I now draw some conclusions from my extended discussion of the Free Will Theorem. I am certainly swayed by it, so I start with the plus side. Firstly, it is pretty much unique in attempting to bridge the gulf between our most up-to-date knowledge from physics and a traditional philosophical question. Secondly, it gives a precise physical setting where we can probe questions relating to abilities and counterfactual scenarios under determinism. This precision is a boon to philosophers, as these debates are notoriously hard to put on a firm physical footing. I have discussed Landsman's interpretation of the result and added my own two cents on its interpretation. Although I have disagreed with his diagnosis of the problem, I think we share the general outlook on the import of the Free Will Theorem: it highlights an unforeseen difficulty for Lewis's Local Miracles Compatibilism. This difficulty resulted in another impasse between PR1, which Landsman and I implicitly endorse, and PR2, which Lewis must resort to, if he is to be safely deny *Freedom*. Other philosophers may want to break this impasse by chasing up the causation question. This may prove fruitful, but also fraught with difficulty as they would need to square causation with non-locality.

On the minus side, the precision of the result is also a drawback. Landsman (2016, p. 1) himself admits it challenges Lewis in "a contrived way via bipartite EPR-type experiments". Lewis could simply bite the bullet and recant Alice's ability to choose settings freely. After all, this is a highly unusual scenario, and our ordinary intuitions are led astray by the weirdness of quantum mechanics.

In the next section I will discuss a second recent challenge to LMC based on physical arguments put forward by Dorr.

4.4 Dorr Against Counterfactual Miracles

The starting point for this section is the claim M we saw in section 2.5.1, which I reissue below:

M If counterfactuals:

a) require possible world semantics;
b) are compatible with determinism;
c) require perfect similarity in the lead up history;

then a contrary-to-fact antecedent can only be true because of a miracle.

Hence, under determinism, counterfactuals can only be true in a non trivial way if we posit miracles. This connection between counterfactuals and miracles is also the starting point for Lewis's Local Miracles Compatibilism. We saw how Lewis could use M to argue that our actions follow in the stream of miracles but do not cause them. And that claim in turn justifies his compatibilism in the face of the Consequence Argument. So by arguing against the need

for miracles in counterfactuals, Dorr is also challenging Lewis' Local Miracles Compatibilism. In the next two sessions I assess the import of this challenge. Dorr's begins by taking issue with requirement c) above:

> vindicating the reliability of our ordinary method of forming counterfactual beliefs does not require taking the strict view that if things had gone differently during an interval *t*, *absolutely all* facts about history before *t* would have been *exactly* as they actually were. It would be enough to say that history before *t* would have proceeded *approximately* as it did in the actual world
>
> (Dorr 2016, p. 252)

Dorr draws a distinction between propositions whose truth that stay true in counterfactual scenarios and those which we can plausibly allow to become false. Everything hangs on getting the right sense of 'approximately' in the above passage. So the basic desideratum is to make sure the everyday historical propositions we hold fixed in practice remain true in any analysis of counterfactuals. These propositions relate to times before the contrary to fact antecedent. The second desideratum is to leave enough room for contrary to fact antecedents, without prejudging the question of determinism. Dorr's ultimate aim is to weaken c) enough to do away with the need for miracles whilst retaining the key idea of Lewis's analysis, namely the similarity ranking of possible worlds.

Lewis was well aware of these desiderata and gave a general implausibility argument as to why they could not be satisfied jointly on physical grounds (1979, p. 45).[51] I will not rehearse Lewis's argument as I will cover a lot of the same ground when I discuss the details of Dorr's proposal. Furthermore, it has been rumoured for a while now that statistical mechanics considerations have some part to play: whether by allowing c) to be weakened (Bennett 2003, p. 219; Wilson 2014, p. 270–271); or whether by breaking with Lewis's analysis to take entropy increase into account explicitly (Elga 2001; Kutach 2002; Loewer 2007). However, to date, the former considerations have stayed on the same level of generality (and rigour) of Lewis's original arguments (uncharitably: where he said implausible they say plausible). The latter analyses are either not as ambitious in their scope or have failed to gain traction. Thus we come to Dorr. What distinguishes his contribution is that he offers a quantitative physical claim to underpin his argument, which he calls the "Independence Conjecture". This claim arises in statistical mechanics.

Dorr actually gives two arguments, and he seems to consider the second one a concrete instance of the first generic argument. In the next two sections I will discuss them in turn.

[51] See also (Bennett 2003, §82) for an overview of the controversy surrounding this claim.

4.5 Dorr's First Argument

I have put Dorr's argument schematically for ease of discussion.[52] It runs as follows (2016, p. 252–254):

P1 "Our best deterministic theories have *continuous dynamics*." (The flow map $\Phi : \mathbb{R} \times X \to X$ is a continuous function.)

P2 "We can regiment Lewis's time-relative notion of similarity between possible worlds using a metric[53] d on X. [...] It is plausible that d is continuous."

P3 "On any reasonable similarity metric d on X, [...] such that $d(p, q)$ is small, and Δt is (say) one second, $d\big(\Phi(\Delta t, p), \Phi(\Delta t, q)\big)$ is some large multiple of $d(p, q)$." (This means the system is chaotic.)

P4 "The prevalence of chaos suggests [...] we should expect the set of worlds that approximately match actuality until t to be quite varied as regards history after t."

$\therefore$ Dynamical considerations support an analysis of counterfactuals without miracles (subject to our desiderata in 4.4).

I am prepared to concede this conclusion once Dorr tells us exactly what 'approximately' means. Put differently: how do we form the d function? In a footnote Dorr says it is enough for his argument that d is a continuous function and $d(x, x) = 0$ for all physical states x, rather than a metric. Moreover, Dorr also need to explain how to move from the similarity of physical states formalized by d to the similarity ranking of worlds discussed by Lewis (in this framework worlds would be time parametrized curves in the space of states X).

Perhaps Dorr thinks we can use the phase space geometry and its metric to define d. Most propositions are true in *sets* of possible worlds, certainly so everyday ones. This may afford enough wiggle room to change the physical state without falsifying these historical propositions.

Harking back to Captain Savitsky, the claim is that there was a state of the world geometrically close to the actual one such that: the B-59 submarine was still in Cuban waters; Julius Caesar was still murdered by Brutus; the Pettakere cave paintings were still made by early humans... The past is different all the way back to the Big Bang, but the vaguish historical propositions which interest us all remain true in the alternative world, even though the B-59 was moving faster; Caesar fell slightly more to the right; the hand paintings are a slightly darker red etc. And most importantly the torpedo *was* launched (and nuclear war did ensue) through some deterministic and lawful divergence.

[52] For compatibility with my earlier discussion, I have changed references to the state space M in the original, to X. Likewise I have changed time intervals x to Δt.

[53] Dorr later weakens this requirement (2016, fn. 18).

Dorr's hope is that all of this can be encoded into a function d which returns numerical values for historical similarity (or dissimilarity). If this function exists and is continuous then Dorr argues we can keep the point-wise dissimilarity between the actual world-curve

$$\alpha\colon [t_0, +\infty) \to X$$

and an alternative world curve

$$w\colon [t_0, +\infty) \to X$$

as small as we like throughout an arbitrary long but finite time interval:

$$\forall \epsilon > 0\ \exists w(t)\ \forall t \in [t_0, 0]\colon d(\alpha(t), w(t)) < \epsilon \tag{4.14}$$

where t_0 and $t = 0$ indicate the first and present instants respectively.

However I worry about how the continuity requirement will be enforced. In particular I worry our judgements of historical similarity may contain singular limits precisely because there may be points where one of our historical propositions suddenly changes truth value. For example: in the actual world I bite into an apple and find half a worm.[54] In an alternative world I bite into the apple and find a quarter of a worm. In another an eighth, then a hundreth etc. These cases are progressively dissimilar from actuality whilst retaining the truth of the historical proposition 'I bit into a worm'. Let us suppose with Dorr they can be assigned a smoothly increasing numerical value for historical dissimilarity. But then I bite into the apple and find no worm at all. And that world must surely lead to a jump in dissimilarity compared to any world with even a little extant worm! So the function cannot be continuous even if it exists.

You may say the counterexample is tendentious; or perhaps the connection with geometrical distance is spurious; or even the vague boundary of macroscopic objects comes to the rescue and blocks singular limits. I acknowledge this worry is not fatal—it cannot be as Dorr has not given the details of d. However, even if he did, then he would leave P1 hostage to fortune. Physics hitherto has used continuum models for spacetime, nature may turn out to be fundamentally discrete. If so, it will be much more difficult to prove as Dorr (*ibid.*, fn. 18) attempts to do, that we can always make the differences throughout the past as numerically small as we like, let alone 'historically' small enough.

This leads me to conclude that in the absence of a detailed explanation of how approximate similarity is to be cashed out, Dorr's first argument is moot. In the second argument he gives a specific meaning to 'approximately'. By doing so he contradicts his claim that the latter argument is a concrete case of the former one. But if my arguments in this section are correct, this actually helps his wider case. I turn to this second argument in the next section.

[54] This example was inspired by (Berry 2002).

4.5.1 Dorr's argument from Statistical Mechanics

The central thrust of Dorr's argument from Statistical Mechanics (SM) is that there are "nomically possible worlds that match actuality macroscopically up to now" (2016, p. 257) where thermodynamical 'miracles' occur and bring about contrary to fact antecedents.[55] Specifically, Dorr argues for the generic physical plausibility[56] of such a situation on quantitative grounds.

I note two things. First, 'approximate' match in the past now acquires a new definite meaning in terms of the SM notions of 'macrostate' and 'macrohistory':

A Two worlds match 'approximately' until now iff they have identical macrohistories.

I will expand on this shortly. Second, thermodynamic miracles are extraordinarily improbable but not impossible events according to the kind of *microscopic* laws discussed until now. In bare outline Dorr's argument is that:

S1 There is a distinction between the microscopic history and the macroscopic history of a system.

S2 Deterministic physical laws only prescribe the microscopic history of a system (given an initial condition), whilst the truth of our historical propositions depends on the macrohistory.

S3 Statistical Mechanics provides a generic example of identical macrohistories diverging at t so that counterfactual antecedents become true.

$\therefore$ SM supports an analysis of counterfactuals without miracles (subject to our desiderata in § 4.4).

For example, the air molecules in a room could all bunch up so as to slam a door shut. That would look like a miracle but is 'only' astronomically improbable from the point of view of classical mechanics. As usual the difficulty is not in showing an instantaneous configuration is possible (just place all the air molecules behind the door with the right speeds) but showing that such a configuration could evolve from a history relevantly like the actual one. Dorr claims we can

> appeal to what I will dub the 'Independence Conjecture', a certain plausible mathematical claim about the behaviour of dynamical maps that plays a central role in statistical mechanics. This conjecture is expressed in terms of the notion of "macrostate": a set of dynamical states that agree on a certain set of statistical quantities, such as the mean temperature, pressure, density and momentum of gas within each cell in some fine-grained lattice.
>
> (*ibid.*, p. 255–256)

Returning to our example, in the alternative world, the room had the same

[55] If he is right then Lewis's worries to the contrary are unwarranted and M falls with them.

[56] The relevant passage is the second paragraph of (Dorr 2016, p. 257), but he then partially retracts this on the following page.

temperature, the same pressure, etc. throughout the past leading up to the moment the door was shut. Thus the macrohistory of this alternative world matches the actual macrohistory. According to Dorr, this macrohistory and the subsequent thermodynamic miracle, are compatible with the laws of classical mechanics as they apply to particles. And we know they are compatible thanks to the 'Independence Conjecture' which is used in SM.

These arguments present considerable difficulties once we look at them in detail. I outline them first and then discuss each in turn in the following subsections. Firstly: how does Statistical Mechanics define macrostates and macrohistories and does this definition warrant S2? Secondly: what is the Independence Conjecture and how is it related to Statistical Mechanics? Thirdly: does the Independence Conjecture support S3?

4.5.2 Macrohistories, macrostates and S2

Suppose we are given a physical system S. For systems that might interest us like tables and people this already involves a leap. The fundamental physics of S does not deal in legs or or mental states, but in atoms or fields or whatever ultimate physics says. Thus we may assign different state spaces to S depending on what level of description we have in mind, with X being the fundamental one. I have already discussed how this leads to the supervenience picture of physicalism. Statistical Mechanics could be seen as an organized, scientific attempt to put the relation between these two levels on a quantitative footing.

A set of macrostates $\mathcal{M}$ for a system is a partition of its state space, i.e. a set of mutually disjoint, jointly exhaustive subsets of X (Butterfield 2012, p. 107). Any given macrostate $M \in \mathcal{M}$ (or cell in the coarse-graining) corresponds to a set of many microstates $x \in X$. From the point of view of the macroscopic observer, all these microstates 'look the same'. If the system is in any of these microstates then it will be found to be in the same macrostate.

A macrohistory of a system is a time-parametrized sequence in the space of macrostates:[57]

$$h : \mathbb{Z} \to \mathcal{M}. \tag{4.15}$$

Properly speaking all macrohistories should range over the same time interval to deserve this name. This interval should be determined by the actual world. If the past is finite and starts at the Big Bang in the actual world then that instant will be $t_0 \in \mathbb{Z}$ and all later times will be $t_i > t_0$. In short a macrohistory is the history of a system from a God's eye perspective. In practice we will often consider special segments of a macrohistory like the finite past $[t_0, 0]$ or the open future $[0, +\infty)$.

Physical quantities like temperature or pressure are examples of macroscopic states. They correspond to many many possible atomic configurations

[57] The time dependence is treated discretely for macrohistories for reasons I will not go into here, see for example (Frigg 2010, § 6.4.3).

(microstates x) in the microscopic description.[58] In this framework the microscopic dynamics are assumed to be deterministic at the outset, whilst, for now, the question is left open for the macroscopic dynamics. In other words, we are not assuming at the outset that a flow map $F : \mathbb{Z} \times \mathcal{M} \to \mathcal{M}$ exists. The macrostates considered in SM are closer in scale to the everyday objects that our usual counterfactuals deal with. Few of us have measured an atom's position or speed, but we regularly measure a room's temperature and act accordingly e.g. put a jumper on. Hence it is plausible that keeping the macrostate fixed will ensure the historical propositions we are interested also stay true. However I note this plausibility rests on a new assumption compared to the supervenience physicalism of earlier sections. Now entities and objects in historical propositions must supervene on the macrostates which in turn supervene on the microstates.[59] This way if there is no change in the macrostates, then there will be no change in the truth value of the historical propositions either. And S2 will follow also.

Finally this assumption also provides a new perspective on the relations between residents in Earman's garden (cf. T3*).

4.5.3 The Independence Conjecture

I start with Dorr's own words (adapted to my notation as before) on the Independence Conjecture:

> To be precise: for any $S \subset X$, let $\Phi(\Delta t, S) = \{q : \exists p \in S(q = \Phi(\Delta t, p))\}$ be the result of evolving all points in S by Δt units of time. Let u be some small (but not too microscopically small) positive real number, representing a unit of time. Where Θ^- is a finite set of ordered pairs of negative multiples of u and macrostates, and Θ^+ is a finite set of ordered pairs of positive multiples of u and macrostates, let $\Xi^- = \bigcap\{\Phi(x, S)|\langle x, S\rangle \in \Theta^-\}$, and $\Xi^+ = \cap\{\Phi(x, S)|\langle x, S\rangle \in \Theta^+\}$. Finally, let M_0 be some macrostate, and for any measurable $S \subset X$, let $P_{M_0}(S)$ equal the volume of $S \cap M_0$ (according to the natural measure) divided by that of M_0. Then the Independence Conjecture says that
>
> $$P_{M_0}(\Xi^- \cap \Xi^+) \approx P_{M_0}(\Xi^-)P_{M_0}(\Xi^+).$$
>
> (2016, fn. 21)

In the above passage Θ^- and Θ^+ are the past and future segments respectively of one or more macrohistories (considered as the ordered pairs that form the sequences). The macrostate at the present time is M_0. It is less clear what the physical meaning of Ξ^- and Ξ^+ is. In particular it seems perverse to evolve each macrostate in the past and future segments by their own time index. This would take past macrostates to subsets of X further back in time, and contrariwise for future macrostates. For a given macrostate M^*, the subset

[58] Foundational work in SM usually assumes classical mechanics is the microscopic theory. This has not gone unchallenged, see (Wallace forthcoming, p. 3).

[59] Frigg elevates this last supervenience relation to a "central assumption of Boltzmannian SM" (Frigg 2010, p. 93).

$X^* = \Phi(\Delta t, M^*)$ includes all the possible microstates the system could be in at a time $t^* + \Delta t$. If the time interval Δt is negative this will be in past compared to M^*, in the future contrariwise. Thus it makes little physical sense to take the intersection of sets of microstates at different times as Ξ^+ and Ξ^- do. I think the most charitable interpretation is that Ξ^- (or Ξ^+) should represent the microstates compatible with a sequence of past (or future) macrostates under the microscopic dynamics. That is, we take each macrostate and wind it forward (or backwards) by the modulus of its time index. This returns the subset of X at time $t = 0$ which each macrostate in the sequence *would have* evolved to (from) under the microdynamics. The intersection of all these sets gives the set of microstates which is compatible with that sequence of past (or future) macrostates. This all amounts to an extra minus sign:

$$\Xi^- := \bigcap \{\Phi(-u, S) \colon \langle u, S\rangle \in \Theta^-\}, \tag{4.16}$$

$$\Xi^+ := \bigcap \{\Phi(-u, S) \colon \langle u, S\rangle \in \Theta^+\}. \tag{4.17}$$

On this reading the Independence Conjecture says that a system being in a microstate compatible with the macropast segment Θ^- is probabilistically independent of it being in a microstate compatible with the macrofuture segment Θ^+, given the system is presently in macrostate M_0.

This Independence Conjecture is a very strong claim about a system. Sometimes it is a justified assumption. Take the case of a gambler who believes a number must come out in the lottery soon because it has not for the last few years. He commits a fallacy because the probability of some future sequence of draws is not affected by the past results. Here the Independence Conjecture holds. But it holds because we are ensuring it holds by implicitly assuming the draw is fair and there is not some hidden microdynamical reason which would prevent certain future sequences from arising after a given past sequence. (Gamblers of the world unite! you have nothing to lose but your microdynamical chains...)

Thus we should be wary of accepting this conjecture too quickly. Dorr argues that we should accept this conjecture as it is a special case of *another* conjecture which, according to Wallace, plays a central role in Statistical Mechanics. Wallace dubs it the Simple Dynamical Conjecture" (SDC). I will examine the conjecture and the nature of the role it plays next.

The Simple Dynamical Conjecture and SM

The Simple Dynamical Conjecture is put forward in a paper where Wallace (forthcoming, p. 2; 8) aims to explain

i) "how the logic of deriving macrophysical irreversibility from microdynamics is supposed to go"

ii) Certain approximation schemes which we use to "make general claims about systems' macrodynamics and to produce closed-form equations for the macrodynamics of specific systems"

iii) "what [Wallace] believes to be *mathematically* going on in these approximation schemes, and what assumptions of a purely technical nature need to be made" setting aside "philosophical and conceptual questions".

I admit at the outset I am sceptical of Wallace's way of proceeding. Whilst i) is a commendable goal, I think it will be difficult to achieve iii) as straightforwardly as he hopes. The reason the foundations of statistical mechanics are as controversial as they are—that cannot be denied!—is in part due to the differing mathematical tools different schools use. I submit this fact is made abundantly clear in two recent reviews of the field (Uffink 2007, p. 923; Frigg 2008, p. 101). These differing mathematical tools reflect differing philosophical assumptions and therefore it is too glib a move to claim the two can be so easily decoupled.

Wallace wants to provide a general framework for topics which range from the Boltzmann Transport Equation to environment-induced decoherence. According to Wallace, these topics share the use of some form of 'coarse graining approximation' in order to derive macroscopic equations—that is, equations which govern the dynamics of the system at the level of macrostates. However, absent any detailed discussion of how these macroscopic results are derived, it is hard to tell whether his framework has indeed picked out a common mathematical core that all the schools above agree on. If not, his approach may just be *one* of many ways of deriving these results. Furthermore, should they all prove to be mathematically sound, then it will be precisely philosophical and conceptual considerations that adjudicate between them.

But let us set aside these worries for the moment and examine the actual content of Wallace's conjecture. We consider a system with probability distribution $\rho(x)$ at t over its state space X. Each point in X evolves according to the deterministic flow map Φ as before. The flow map Φ is determined in principle by classical mechanics. It also induces a corresponding 'Liouville' evolution map

$$L\colon (\rho, \Delta t) \mapsto L(\Delta t)\cdot\rho, \tag{4.18}$$

which takes a distribution at ρ at t and evolves it to ρ' at $t+\Delta t$. I stress this is all still part of the microscopic dynamics of the system as predicted by classical mechanics.
We then choose a coarse graining rule on the space of distributions

$$\mathcal{C}\colon \rho \mapsto \mathcal{C}(\rho) = \rho'$$

subject to some plausible requirements.[60] Wallace discusses several specific maps which are used in practice, but in keeping with his aims he does not try to justify why. The coarse graining map also induces a time evolution operator by:

$$L_{\mathcal{C}}\colon (\rho, \Delta t) \mapsto L_{\mathcal{C}}(\Delta t)\cdot\rho. \tag{4.19}$$

As I read him, Wallace is arguing that the choice of coarse-graining combined with L gives us an analogous evolution operator $L_{\mathcal{C}}$ in a way which varies with

[60] The maps must be projections $\mathcal{C}^2 = \mathcal{C}$ which preserve the total probability of a given macrostate and must commute with time-reversal (Wallace forthcoming, p. 8–9).

the details of the application at hand. The main idea is that the coarse-graining interrupts the natural microscopic evolution given by L periodically with very small period. Once again it is somewhat hard to tell due to the lack of detailed examples. But let us grant the assumption that the choice of coarse-graining map also induces a coarse-grained time evolution operator. This gives the macroscopic dynamics of the system.

In this framework we are primarily interested in assigning probabilities to macrohistories. These assignments must agree with those implicit in the distribution ρ. Wallace achieves this as follows.

Suppose we know the system started out in point x_0. The map Φ then determines all its other microstates thereon. Suppose the microscopic evolution intersects the macrohistory h at all times:

$$\forall t_i \; \Phi(t_i - t_0, x_0) \in M_i = h(t_i). \tag{4.20}$$

If so then h occurs with certainty:

$$P(h|x_0) = 1. \tag{4.21}$$

Analogously, suppose the system started in $\rho_0(x)$. The probability that system is in macrostate M_0 at t_0, given this information is

$$P\big((M_0, t_0)|\rho_0\big) = \int_{M_0} \rho_0(x) = \int_X R(M_0) \cdot \rho_0, \tag{4.22}$$

where $R(M_0) \cdot \rho_0$ indicates the distribution obtained by restricting ρ_0 to $M_0 \in \mathcal{M}$. We could evolve ρ_0 under the Liouville dynamics to get the distribution at t_1

$$\rho_1(x) = L(t_1 - t_0) \cdot \rho_0(x). \tag{4.23}$$

Now the probability that the system went through macrostates M at t_0, and M_1 at t_1 given it started in ρ_0 is

$$P\big((M_1, t_1) \wedge (M_0, t_0)|\rho\big) = P\big(M_1|R(\rho, M_0)\big) = \int_X L(t_1 - t_0) \cdot R(M_0) \cdot \rho_0. \tag{4.24}$$

This conditionalizing procedure can be repeated to generate two history operators which evolve a given initial distribution $\rho_0(x)$ through a macrohistory $h\colon [t_0, t_f] \to \mathcal{M}$ to some final distribution $\rho_f(x)$:

$$H(h) \cdot \rho_0 := R(h(t_f)) \cdot L(t_f - t_{f-1}) \cdots L(t_2 - t_1) \cdot R(h(t_1)) \cdot \rho_0, \tag{4.25}$$

$$H_{\mathcal{C}}(h) \cdot \rho_0 := R(h(t_f)) \cdot L_{\mathcal{C}}(t_f - t_{f-1}) \cdots L_{\mathcal{C}}(t_2 - t_1) \cdot R(h(t_1)) \cdot \rho_0. \tag{4.26}$$

Thus we can assign probabilities to whole macrohistories using these history operators:

$$P(h|\rho) = \int_X H(h) \cdot \rho, \tag{4.27}$$

$$P_{\mathcal{C}}(h|\rho) = \int_X H_{\mathcal{C}}(h) \cdot \rho. \tag{4.28}$$

Wallace states the condition which a system must obey in order that the microscopic dynamics and the macroscopic dynamics obtained by coarse-graining agree:

$$\forall h \quad \mathcal{C} \circ H(h) \cdot \rho = H_{\mathcal{C}}(h) \cdot \rho. \tag{4.29}$$

Note this also ensures the probabilities $P(h)$ and $P_{\mathcal{C}}(h)$ agree thanks to one of the conditions on the coarse-graining (see fn. 60).

Finally the Simple Dynamical Conjecture is a conjecture about which initial distributions obey this condition given a system S and a coarse-graining $\mathcal{C}$. It says that

> Any distribution [representing S] whose structure is at all simple is forward predictable by $\mathcal{C}$ [i.e. obeys equation 4.29 for $\mathcal{C}$]; any distribution *not* so predictable is highly complicated and as such is not specifiable in any simple way *except* by stipulating that it is generated via evolving some other distribution in time [...].
>
> (Wallace forthcoming, p. 19)

The Independence Conjecture and SM

I will now try to answer the question posed at the end of section 4.5.1. I have already stated Dorr's Independence Conjecture and Wallace's Simple Dynamical Conjecture and explained their technical framework. What then is the relation between the Independence Conjecture and Statistical Mechanics? According to Dorr

> There is considerable empirical evidence for the Independence Conjecture. As Wallace (*ibid.*) persuasively shows, it—or rather a more general conjecture of which it is a special case—plays a ubiquitous role in statistical-mechanical "derivations" of equations governing macroquantities (such as the Boltzmann equation); and many of these equations have proved to be fantastically accurate.
>
> (Dorr 2016, p. 256–257)

So, in outline: certain SM equations are empirically successful; the SDC is crucial in deriving these SM equations; the IC is a special case of the SDC.

This is my critical re-reading of this outline: idem; according to Wallace the SDC is assumed when deriving these SM equations relative to some *choice* of coarse-graining; according to Dorr the IC is a special case of the SDC assuming—at the very least—the system had a 'simple' initial distribution.

Let's grant, for the sake of argument, the first point and Dorr's claim that the IC is a special case of the SDC.[61] There are two difficulties inherent in his appeal to Wallace's conjecture. Firstly, how exactly is the choice of coarse-graining to be made? Here I find it hard to ignore the "philosophical and conceptual worries." Do the details of the system necessitate one particular map $\mathcal{C}$? If so, how? As already noted, Wallace does not say.

[61] Dorr has kindly provided me with a proof of this fact, given he did not give one in the original paper. I decided not to include it here for reasons that will become apparent.

If not, perhaps we should just stick with the choice practising physicists make and report in articles and textbooks. But I resist the idea that the justification should just be 'we choose whatever coarse-graining we have found yields the empirically adequate equations'. In that case, all we are describing is a complicated and informed guessing procedure. We are certainly not deriving "macrophysical irreversibility from the microdynamics" in any meaningful way, as per Wallace's stated aim i).

Secondly, how do we determine whether a system obeys the SDC, i.e. whether its initial distribution was 'simple'? I propose to do as Dorr does and ask Wallace who admits:

> the notion of "simplicity" is hard to pin down precisely, and I will make no attempt to do so here. (If desired the Simple Dynamical Conjecture can be taken as a family of conjectures, one for each reasonable precisification of "simple".) [...]
> Are individual states (that is, classical single-system states or quantum pure states) Simple? It depends on the state in question. Most classical or quantum states are not Simple at all: they require a great deal of information to specify. But there are exceptions: some product states in quantum mechanics will be easily specifiable, for instance; so would states of a classical gas where all the particles are at rest at the points of a lattice. This in turn suggests that the Simple Dynamical Conjecture may well fail in certain classical systems *(specifically, those whose macrodynamics is in general indeterministic)*.
>
> (Wallace forthcoming, p. 19, my italics)

If Wallace is right about his conjecture, this is a problem for Dorr. It undermines premise S3 of his argument from SM, which I recall below:

S3 Statistical Mechanics provides a generic example of identical macrohistories diverging at t so that counterfactual antecedents become true.

This premise requires the system obeys indeterministic macrodynamics so as to ensure the worlds diverge after an identical macropast. And that is precisely what Wallace thinks might stop a system from obeying the SDC.[62]

In conclusion I think SM does not lend unqualified support to Dorr's basic contention that miracles are not needed to bring about contrary to fact antecedents. Rather, in my critical reading of Dorr's argument, I have pointed out where the difficulties arise in establishing this claim. In particular, I have pointed out several problematic methodological assumptions Wallace makes together with the inherent vagueness of the SDC conjecture he ends up with. All in all, I think this significantly weakens Dorr's challenge to Lewis miraculous analysis of counterfactuals.

[62] I also note that Butterfield (2012, p. 108) makes a very similar point with regards to macroscopically indeterministic systems. In these systems "micro and macro dynamics do not mesh". For the "micro-dynamics [to] induce a deterministic macro-dynamics" would require "in mathematical jargon: [that] coarse graining and time-evolution commute". Which is precisely what Wallace's equation 4.29 requires. So it is not just Wallace who thinks indeterministic macrodynamics are a problem for the SDC.

In the next section, I examine the final question posed at the end of § 4.5.1. This leads me to discuss an alternative SM school which at the very least provides a clearer foundation for Dorr's Independence Conjecture.

4.5.4 From independent probabilities to existence

Does the Independence Conjecture support S3? Dorr's own view is that

> claims of probabilistic independence yielded by the [Independence] conjecture are much stronger than the mere existence claims that we are concerned with—for example, the claim that there is a nomically possible world whose history approximately matches actuality until t at which the air subsequently squeezes itself into the corner.
>
> (Dorr 2016, p. 257)

I actually agree with Dorr's conclusion but perhaps not for the same reasons. Specifically I do not think it's the independence claim which is doing any of the philosophical work, as I already hinted with the case of the gambler.

Mathematical probability theory as axiomatised by Kolmogorov is based on the notion of probability spaces (Ω, S, P). The set Ω is usually glossed as the set of possible outcomes (or elementary events). In doing so we leave austere mathematics behind and venture into the realm of interpretation. When Dorr claims two segments of a macrohistory can be assigned probabilities he implicitly claims they belong to Ω. Otherwise by definition we are not using standard probabilities. But this poses a problem: anything in Ω is 'possible' according to the standard gloss. So no wonder the Independence Conjecture implies certain worlds are possible! By assuming probabilities apply in the first place he begs the question of the existence of the aforementioned worlds (from the point of view of the laws).

Perhaps there are various notions of possibility at play. Perhaps Dorr does not beg the question because Ω-possible is not the same as nomically possible. We could choose a weak criterion for Ω membership, for example logical possibility. (We certainly cannot use a stronger one on pain of begging the question as before.) But then we have let miracles in again through the back door by allowing miraculous possibilities into Ω. Thus I doubt Dorr would want to resort to such measures.

I think Dorr's best option is to try to reject the first charge by arguing physics tells us there really is stochasticity on the macro level. In fact, this is my own interpretation of his argument on balance. Indeed Dorr's own gloss of the Independence Conjecture is that

> "the macropresent screens off the macrofuture from the macropast": in the probability distribution that we get by restricting to a particular macrostate, facts about *future* macrostates are, approximately, probabilistically independent of facts about *past* macrostates.
>
> (*ibid.*, p. 256)

I propose we take this gloss literally and take it as a starting point. This brings

us to the "stochastic dynamics" school of Statistical Mechanics. The basic postulates of this approach to SM are outlined particularly clearly in (Penrose 1970, p. ix).

The fundamental posit of this school is that whilst the microscopic evolution is deterministic the macroscopic evolution is described by a stochastic process (Uffink 2007, p. 1038). In this approach the macrostates becomes random variables, macrohistories and their segments are stochastic processes and they are assigned probabilities much as before. The key additional assumption is that these stochastic processes obey the Markov property (*ibid.*, p. 1044):

$$\begin{aligned} P(M_n, t_n; \ldots; M_1, t_1 | M_i, t_i) = \\ P(M_n, t_n; \ldots; M_{i+1}, t_{i+1} | M_i, t_i) \cdot P(M_{i-1}, t_{i-1}; \ldots; M_1, t_1), \end{aligned} \tag{4.30}$$

for all $n = 1, 2, \ldots$ and all $1 \leq i \leq n$.

This, I submit, is the content of Dorr's Independence Conjecture up to notation.[63] As is Uffink's gloss of it:"the future and past are independent if one conditionalizes on the present" (*ibid.*, p. 1044).

Finally I will not try to settle whether Wallace's approach is part of this wider stochastic dynamics school. I note members of this school take a hard-nosed approach to justifying coarse-graining by appealing to its phenomenological success (see (*ibid.*, §7.5)). Nonetheless, if they are different denominations of the same basic faith, I prefer the latter for making its tenets clearer. We are explicitly assuming from the beginning that nature behaves like a series of coin flips on the macroscopic level. Thus by definition there are no nomically impossible sequences of coin-flips, which is hardly surprising.

4.6 Moral of Dorr's Arguments

I have discussed Dorr's arguments from continuous dynamics and from Statistical Mechanics which challenge the need for miracles in counterfactuals and thus undercut Local Miracle Compatibilism.

I hope to have convinced you both these arguments establish considerably less than initially advertised. In my extended critique of the latter I have highlighted two main issues.

Firstly, as presented by Dorr, the argument relied uncritically on a conjecture put forward by Wallace. My main complaint was that Wallace's coarse-graining conjecture cannot justify the substantive conclusions Dorr requires, given Wallace's methodology in reaching it.

Secondly, as I re-interpreted it, the Independence Conjecture is really just a stochasticity postulate at the macroscopic level. This postulate does play a role in SM in the stochastic dynamics research school, but its philosophical justification has remained elusive.

All in all, Dorr's physical arguments do not threaten Lewis. Nor do they alter the stalemate between him and van Inwagen. But I draw a positive conclusion

[63] I have changed the y_i values of random variables in Uffink's original equation to my M_i and eliminated the subscripts on the probabilities which do not concern us here.

from Dorr's ideas. The layered relation between fundamental physical states, statistical macrostates and historical propositions gives us a new perspective on Earman's garden (see the end of § 2.1).

Dorr's idea that determinism at the fundamental microscopic level need not entail determinism at the macrostate level, nor for that matter determinism (or unavoidability) at the level of commonplace propositions, is worth exploring further. In the next chapter I will speculate as to where that might take us before bowing out with my conclusions.

5 Conclusions

Lewis's compatibilism without actual miracles would be an attractive theory. Here I speculate how that might be achieved drawing upon the lessons learned in the previous chapters.

I say speculate because I will not provide a full-blown defence of what is to follow. My justification can only be the title of this chapter and the fact that many of these ideas have recently been aired by List (2014, 2015, 2016) and List and Pivato (2015), to whom I refer you for the details. Perhaps the moral of this chapter should be that I would have started here had I known what I now know! What I wish to convey is that this would be the way to go if you are unwilling to accept miracles, but not yet ready to embrace semi-compatibilism or even incompatibilism.[64]

5.1 Compatibilism without Actual Miracles

I start by updating the garden of forking paths model of agential choice by taking the existence of different levels of description into account. We should now imagine the forking paths reveal microscopic structure upon closer inspection. Each path is woven out of a bundle of individual filaments which never branch or merge as they extend across the garden. Two filaments may start together but end up in radically different parts of the garden if their paths should there take them. This image has been used by Loewer (2007, fig. 11.1) and List (2014, figs. 1–2) to argue for a compatibilist thesis as follows.

List has advocated a modal reading of ability[65] according to which Alice is able to do otherwise iff it is *agentially possible* for her to do otherwise. Determinism may be true at the microscopic level as we have been assuming all along, but this does not imply determinism at the level of actions. Thus Dorr's mistake was only to have looked in the wrong place, namely SM, for evidence of indeterministic higher level behaviour.

Most importantly, we could have compatibilism without miracles, by finessing the 'may be true' requirement. The FWT is true at the level of agents, while Det is true at the level of fundamental physics. There is nothing inherent to supervenience physicalism which necessitates meshing dynamics (see fn. 62); our experience and the special sciences confirm it; the burden of proof

[64] Assuming you did not make it thus far without being a physicalist of course!

[65] A reading which treat the propositional ability operator as an operator of suitably restricted possibilities, see (Maier 2014, § 4.1ff.)

is on the reductive physicalists to show otherwise. Let not van Inwagen join together, what in Earman's garden was put asunder!

List expresses this more seriously in (2015) were he argues it is the very language of FWT and Det which prevents the Consequence Argument from gaining traction. Premises which "combine propositions asserting fundamental physical facts with operators capturing agential abilities" are far from innocuous. At worst they are not even wrong as they "illicitly [mix] fundamental-physics and agency talk" (*ibid.*, p. 4). It is our tendency, as philosophers, to abstract away from the details that has led us into the error of thinking they were meaningful.

This is all very well, but for this line of argument to even get off the ground, language must surely reflect ontology. There must be genuine chanciness at the agential level. And this is precisely why List has argued for emergent chance in (2015).

If like me, you are not yet ready to swallow non-reductive physicalism and its ontology whole, here is a more cautious variant of this approach. Start, at the very beginning, by disregarding van Inwagen's well-meaning advice and define free will in terms of what is morally relevant. Thus 'Alice acted freely' if

FWTM It was possible for Alice to do otherwise with respect to the coarse-grained past our moral discourse induces.

The claim is that there some wiggle room in our practice of praising, blaming and holding each other to account with respect to the exact physical specification of what went on.

Jack is still a culpable murderer regardless of exactly how fast his hand was travelling when he hit his victim. Conversely the brain tumour patient would have acted just as he did had the brain tumour been in a slightly different spot yet nonetheless we do not hold him responsible. Perhaps because we think the murderer would still have been the same blackguard for morally indistinguishable but physically different pasts, whilst the patient would have genuinely been a different person had he not developed the tumour. Thus we blame the person, in one case but not the other, because we feel it is the person that made the moral difference not the physical state.

If determinism is true, no one can help the physical state they are in, nor can they choose the physical past it descends from. But we can ask whether a certain sequence of actions, or an agent's choices in life, described with only as much detail as is necessary to determine moral responsibility in each case, is compatible with multiple physical outcomes at this same level.

There is still plenty of scope for messy, involved considerations of whether external factors trumped personal agency or vice versa—he could not help a violent childhood; he chose to take those tumour inducing drugs.

This would be compatibilism for the people, by the people, as the truth of FWTM is sensitive to those selfsame considerations by which we would judge ourselves free—or not—in practice. I think it is an interesting starting point.

6 Bibliography

Beebee, Helen (2003).
"Local Miracle Compatibilism".
In: *Noûs* 37.2, pp. 258–277.
DOI: 10.1111/1468-0068.00438
(Cit. on p. 22).

Bennett, Jonathan (2003).
A Philosophical Guide to Conditionals.
Oxford University Press,
P. 387.
DOI: 10.1093/0199258872.001.0001
(Cit. on pp. 20, 21, 34, 45).

Berry, Michael (2002).
"Singular limits".
In: *Physics Today 55.5*, pp. 10–11.
DOI: http://dx.doi.org/10.1063/1.1485555
(Cit. on p. 47).

Burns, JM and Swerdlow, RH (2003).
"Right orbitofrontal tumor with pedophilia symptom and constructional apraxia sign".
In: *Archives of Neurology* 60.3, pp. 437–440.
DOI: 10.1001/archneur.60.3.437
(Cit. on p. 9).

Butterfield, Jeremy (2012).
"Laws, causation and dynamics at different levels".
In: *Interface Focus* 2, pp. 101–114.
DOI: 10.1098/rsfs.2011.0052
(Cit. on pp. 33, 49, *55*).

Byrne, Ruth M. J. and McEleney, Alice (2000).
"Counterfactual thinking about actions and failures to act".
In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26.5, pp. 1318–1331.
DOI: 10.1037/0278-7393.26.5.1318
(Cit. on p. 20).

Cator, Eric and Landsman, Klaas (2014).
"Constraints on Determinism: Bell Versus Conway–Kochen".
In: *Foundations of Physics* 44.7, pp. 781–791.
DOI: 10.1007/s10701-014-9815-z
(Cit. on pp. 27, 29, 30).
Conway, John and Kochen, Simon (2006).
"The Free Will Theorem".
In: *Foundations of Physics* 36.10, pp. 1441–1473.
DOI: 10.1007/s10701-006-9068-6
(Cit. on p. 27).
— (2009).
"The strong free will theorem".
In: *Notices of the AMS* 56.2, pp. 226–232.
URL: http://www.ams.org/notices/200902/rtx090200226p.pdf (visited on 17/10/2016)
(Cit. on pp. 27, 28).
Crane, Tim and Mellor, D. H. (1990).
"There is No Question of Physicalism".
In: *Mind* 99.394, pp. 185–206.
DOI: 10.1093/mind/XCIX.394.185
(Cit. on p. 33).
De Mola, Davide (2016).
"Caption Contest #539".
In: *The New Yorker*. (sadly unpublished).
URL: http://contest.newyorker.com/CaptionContest.aspx?tab=archive&id=539 (visited on 26/10/2016)
(Cit. on p. 73).
Dorr, Cian (2016).
"Against Counterfactual Miracles".
In: *Philosophical Review* 125.2, pp. 241–286.
DOI: 10.1215/00318108-3453187
(Cit. on pp. 6, 11, 27, 45–48, 50, 54, 56).
Drábik, Peter (2007).
"On Disjunction in Modal Logics".
MA thesis. Comenius University in Bratislava,
P. 56.
URL: http://diplomovka.sme.sk/praca/3002/on-disjunction-in-modal-logics.php?g{_}for=Drabik
(Cit. on p. 24).
Earman, John (1986).
A primer on determinism.
Dordrecht: D. Reidel Publishing Company,
P. 273
(Cit. on pp. 6, 7, 10, 11, 21, 28).

— (1989).
World Enough and Space-Time: Absolute versus Relational Theories of Space and Time.
Cambridge, MA: MIT Press,
P. 233
(Cit. on p. 10).
— (2007).
"Aspects of Determinism in Modern Physics".
In: *Philosophy of Physics: Part B*.
Ed. by Jeremy Butterfield and John Earman.
Amsterdam: North Holland,
Pp. 1369–1434
(Cit. on p. 11).
Elga, Adam (2001).
"Statistical Mechanics and the Asymmetry of Counterfactual Dependence".
In: *Philosophy of Science* 68.3, S313–S324.
DOI: 10.1086/392918
(Cit. on p. 45).
Fischer, John Martin (2012a).
Deep Control: Essays on Free Will and Value.
Oxford University Press,
P. 244.
DOI: 10.1093/acprof:osobl/9780199742981.001.0001
(Cit. on p. 6).
— (2012b).
"Semicompatibilism and Its Rivals".
In: *The Journal of Ethics* 16.2, pp. 117–143.
DOI: 10.1007/s10892-012-9123-9
(Cit. on p. 6).
Frankfurt, Harry G. (1969).
"Alternate Possibilities and Moral Responsibility".
In: *The Journal of Philosophy* 66.23, pp. 829–839.
DOI: 10.2307/2023833
(Cit. on p. 6).
Frigg, Roman (2008).
"A Field Guide to Recent Work on the Foundations of Statistical Mechanics".
In: *The Ashgate Companion to Contemporary Philosophy of Physics*.
Ed. by D. Rickles.
London: Ashgate Pub. Limited,
Pp. 99–196
(Cit. on p. 52).

Frigg, Roman (2010).
"Probability in Boltzmannian statistical mechanics".
In: *Time, Chance, and Reduction: Philosophical Aspects of Statistical Mechanics*.
Ed. by G. Ernst and A. Hüttemann.
Cambridge University Press,
Pp. 92–118
(Cit. on pp. 49, 50).
Graham, Peter A. (2008).
"A defense of local miracle compatibilism".
In: *Philosophical Studies* 140.1, pp. 65–82.
DOI: 10.1007/s11098-008-9226-0
(Cit. on p. 22).
Hempel, Carl G. (2001).
"Reduction: Ontological and Linguistic Facets".
In: *The philosophy of Carl G. Hempel: studies in science, explanation, and rationality*.
Ed. by James H. Fetzer.
New York: Oxford University Press, Inc.
Chap. 9, pp. 189–207
(Cit. on p. 33).
Horgan, Terence (1985).
"Compatibilism and the consequence argument".
In: *Philosophical Studies* 47.3, pp. 339–356.
DOI: 10.1007/BF00355208
(Cit. on pp. 22, 25).
Huemer, Michael (2000).
"Van Inwagen's Consequence Argument".
In: *The Philosophical Review* 109.4, pp. 525–544.
DOI: 10.2307/2693623
(Cit. on p. 14).
Hughes, G. E. and Cresswell, M. J. (1998).
A New Introduction to Modal Logic.
London: Routledge,
P. 421
(Cit. on p. 24).
van Inwagen, Peter (1975).
"The incompatibility of free will and determinism".
In: *Philosophical Studies* 27.3, pp. 185–199.
DOI: 10.1007/BF01624156
(Cit. on pp. 13, 16).
— (1983).
An Essay on Free Will.
Oxford University Press,
P. 248
(Cit. on pp. 6, 9, 11–15, 23, 25, 32, 77).

— (2004).
"Freedom to Break the Laws".
In: *Midwest Studies in Philosophy* 28.1, pp. 334–350.
DOI: 10.1111/j.1475-4975.2004.00099.x
(Cit. on pp. 12–14, 16, 22).
— (2005).
"Free Will Remains a Mystery".
In: *The Oxford Handbook of Free Will*.
Ed. by Robert Kane.
1st ed.
Oxford University Press, Inc.
Chap. 7, pp. 158–177.
DOI: 10.1093/oxfordhb/9780195178548.003.0007
(Cit. on p. 26).
— (2008).
"How to Think about the Problem of Free Will".
In: *The Journal of Ethics* 12.3-4, pp. 327–341.
DOI: 10.1007/s10892-008-9038-7
(Cit. on pp. 3, 5, 12, 14).
— (2015).
"Some Thoughts on An Essay on Free Will".
In: *The Harvard Review of Philosophy* 22, pp. 16–30.
DOI: 10.5840/harvardreview2015224
(Cit. on pp. 4, 14).
Jackson, Frank (2000).
From Metaphysics to Ethics.
Oxford University Press,
P. 174.
DOI: 10.1093/0198250614.001.0001
(Cit. on p. 3).
Jennings, R. E. (1995).
The Genealogy of Disjunction.
Oxford University Press,
P. 344.
DOI: 10.1093/acprof:oso/9780195075243.001.0001
(Cit. on p. 24).
Kapitan, Tomis (2011).
"A Compatibilist Reply to the Consequence Argument".
In: *The Oxford Handbook of Free Will*.
Ed. by Robert Kane.
2nd ed.
New York: Oxford University Press, Inc.
Chap. 7, pp. 131–150.
DOI: 10.1093/oxfordhb/9780195399691.003.0007
(Cit. on pp. 15–17).

Knobe, Joshua (2014).
"Free Will and the Scientific Vision".
In: *Current Controversies in Experimental Philosophy*.
Ed. by Edouard Machery and Elizabeth O'Neill.
New York: Routledge.
Chap. 5, pp. 69–85
(Cit. on p. 8).
Kutach, Douglas N. (2002).
"The Entropy Theory of Counterfactuals".
In: *Philosophy of Science* 69.1, pp. 82–104.
DOI: 10.1086/338942
(Cit. on p. 45).
Landsman, Klaas (2016).
On the notion of free will in the Free Will Theorem.
submitted to *Studies in History and Philosophy of Modern Physics*, 17 October 2016 version.
URL: http://www.math.ru.nl/~landsman/FWTLewisv7.pdf (visited on 03/11/2016)
(Cit. on pp. 6, 27–31, 33, 35–37, 44).
Leifer, Matthew (2014).
"Is the Quantum State Real? An Extended Review of Ψ-ontology Theorems".
In: *Quanta* 3.1, pp. 67–155.
DOI: 10.12743/quanta.v3i1.22
(Cit. on p. 32).
Lewis, David K. (1973).
"Causation".
In: *The Journal of Philosophy* 70.17, pp. 556–567.
DOI: 10.2307/2025310
(Cit. on p. 19).
— (1979).
"Counterfactual Dependence and Time's Arrow".
In: *Noûs* 13.4, pp. 455–476.
DOI: 10.2307/2215339
(Cit. on pp. 18, 20, 45).
— (1981).
"Are we free to break the laws?"
In: *Theoria* 47.3, pp. 113–121.
DOI: 10.1111/j.1755-2567.1981.tb00473.x
(Cit. on pp. 9, 11, 12, 14, 17, 18, 21, 22, 36).
— (1983a).
"New work for a theory of universals".
In: *Australasian Journal of Philosophy* 61.4, pp. 343–377.
DOI: 10.1080/00048408312341131
(Cit. on p. 33).

— (1983b).
Philosophical Papers Volume I.
Preface.
Oxford University Press,
P. 285.
DOI: 10.1093/0195032047.001.0001
(Cit. on p. 6).
— (1986).
On the Plurality of Worlds.
Oxford: Blackwell,
P. 276
(Cit. on pp. 21, 39, 40).
— (1987a).
"A Subjectivist's Guide to Objective Chance".
In: *Philosophical Papers Volume II*.
Oxford University Press,
Pp. 83–113.
DOI: 10.1093/0195036468.003.0004
(Cit. on p. 32).
— (1987b).
"Causal Explanation".
In: *Philosophical Papers Volume II*.
Oxford University Press,
Pp. 214–240.
DOI: 10.1093/0195036468.003.0007
(Cit. on p. 17).
— (1987c).
Philosophical Papers Volume II.
Preface.
Oxford University Press,
P. 366.
DOI: 10.1093/0195036468.001.0001
(Cit. on p. 28).
— (1989).
"Dispositional Theories of Value II".
In: *Proceedings of the Aristotelian Society, Supplementary Volumes* 63, pp. 113–137.
URL: http://www.jstor.org/stable/4106918 (visited on 26/10/2016)
(Cit. on p. 35).
— (2001).
Counterfactuals.
2nd ed.
Oxford: Blackwell Publishers,
P. 156
(Cit. on pp. 19–21).

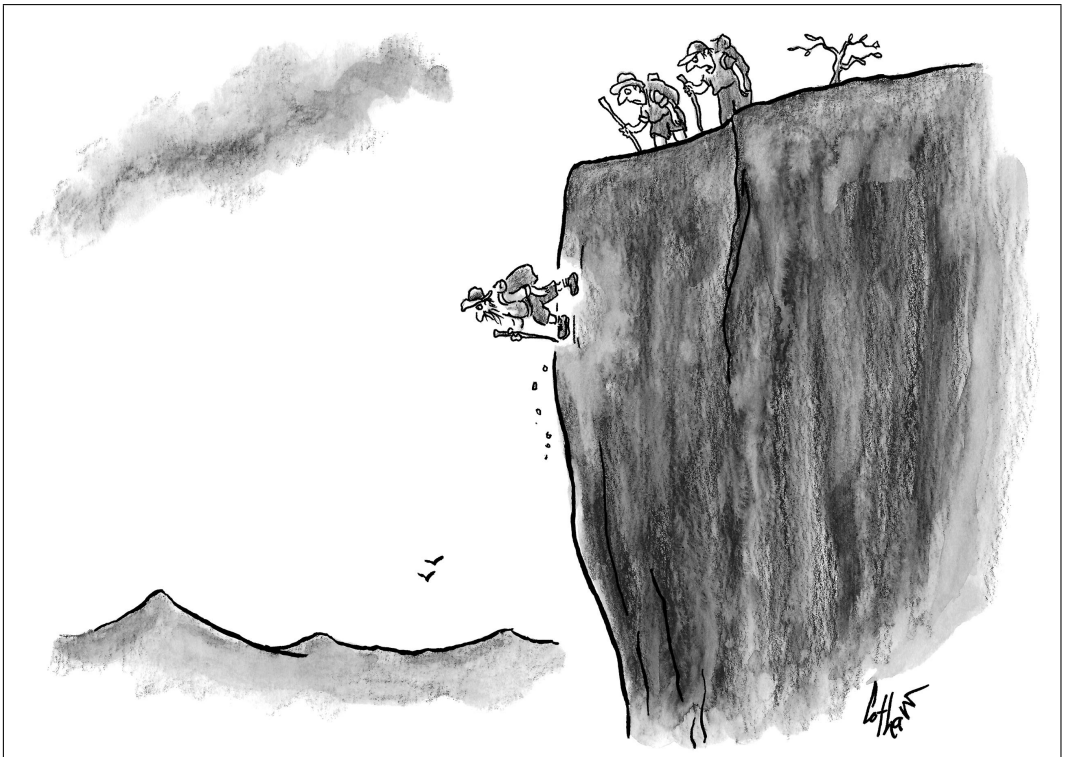
Lewis, Stephanie R. (2015a).
"Bibliography of the Work of David Lewis".
In: *A Companion to David Lewis*.
Ed. by Barry Loewer and Jonathan Schaffer.
Oxford: John Wiley & Sons, Ltd,
Pp. 562–571.
DOI: 10.1002/9781118398593.biblio
(Cit. on p. 14).
— (2015b).
"Where (in Logical Space) Is God?"
In: *A Companion to David Lewis*.
Ed. by Barry Loewer and Jonathan Schaffer.
Oxford: John Wiley & Sons, Ltd.
Chap. 13, pp. 206–219.
DOI: 10.1002/9781118398593.ch13
(Cit. on pp. 14, 21).
List, Christian (2014).
"Free Will, Determinism, and the Possibility of Doing Otherwise".
In: *Noûs* 48.1, pp. 156–178.
DOI: 10.1111/nous.12019
(Cit. on p. 59).
— (2015).
What's wrong with the consequence argument: In defence of compatibilist libertarianism.
PhilSci-Archive preprint, 23 Sep 2015 version
(Cit. on pp. 59, 60).
— (2016).
Levels: descriptive, explanatory, and ontological.
PhilSci-Archive preprint, March-April 2016 version.
URL: http://philsci-archive.pitt.edu/12040/ (visited on 25/10/2016)
(Cit. on pp. 34, 59).
List, Christian and Pivato, Marcus (2015).
"Emergent Chance".
In: *Philosophical Review* 124.1, pp. 119–152.
DOI: 10.1215/00318108-2812670
(Cit. on pp. 59, 60).
Loewer, Barry (2007).
"Counterfactuals and the Second Law".
In: *Causation, physics, and the constitution of reality : Russell's republic revisited*.
Ed. by Huw Price and Richard Corry.
Oxford: Clarendon Press.
Chap. 11, pp. 293–326
(Cit. on pp. 45, 59).

Maier, John (2014).
"Abilities".
In: *The Stanford Encyclopedia of Philosophy*.
Ed. by Edward N. Zalta.
Fall 2014.
URL: https://plato.stanford.edu/archives/fall2014/entries/abilities/
(visited on 03/12/2016)
(Cit. on p. 59).
McKay, Thomas J. and Johnson, David (1996).
"A Reconsideration of an Argument against Compatibilism".
In: *Philosophical Topics* 24.2, pp. 113–122.
DOI: 10.5840/philtopics199624219
(Cit. on p. 26).
Mele, Alfred (2003).
"Agents' Abilities".
In: *Noûs* 37.3, pp. 447–470.
DOI: 10.1111/1468-0068.00446
(Cit. on p. 42).
Moore, G. E. (2005).
"The Nature of Moral Philosophy".
In: *Ethics*.
Oxford University Press,
P. 140.
DOI: 10.1093/0199272018.001.0001
(Cit. on p. 5).
Nichols, Shaun (2011).
"Experimental philosophy and the problem of free will."
In: *Science* 331.6023, pp. 1401–3.
DOI: 10.1126/science.1192931
(Cit. on p. 5).
Oakley, Shane (2006).
"Defending Lewis's Local Miracle Compatibilism".
In: *Philosophical Studies* 130.2, pp. 337–349.
DOI: 10.1007/s11098-004-4677-4
(Cit. on p. 22).
O'Grady, Jane (2001).
Obituary: David Lewis.
The Guardian.
URL: https://www.theguardian.com/news/2001/oct/23/guardianobituaries.books (visited on 10/10/2016)
(Cit. on p. 14).

Oxford University Press (2016).
"freely, adv."
In: *Oxford English Dictionary Online*.
Oxford University Press.
URL: http://www.oed.com/view/Entry/74414?rskey=iAMxwf{\&}result=2{\&}isAdvanced=false (visited on 10/10/2016)
(Cit. on p. 7).
Pendergraft, Garrett (2011).
"The explanatory power of local miracle compatibilism".
In: *Philosophical Studies* 156.2, pp. 249–266.
DOI: 10.1007/s11098-010-9594-0
(Cit. on p. 22).
Penrose, O. (1970).
Foundations of Statistical Mechanics: A Deductive Treatment.
London: Pergamon Press,
P. 260.
DOI: 10.1016/B978-0-08-013314-0.50001-7
(Cit. on p. 57).
Quine, W. V. (1968).
"Propositional Objects".
In: *Crítica: Revista Hispanoamericana de Filosofía* 2.5, pp. 3–29.
URL: http://www.jstor.org/stable/40103900 (visited on 02/11/2016)
(Cit. on p. 40).
Russell, Bertrand (1912).
"On the Notion of Cause".
In: *Proceedings of the Aristotelian Society* 13, pp. 1–26.
URL: http://www.jstor.org/stable/4543833 (visited on 30/10/2016)
(Cit. on p. 28).
Sarkissian, Hagop, Chatterjee, Amita, De Brigard, Felipe, Knobe, Joshua, Nichols, Shaun and Sirker, Smita (2010).
"Is Belief in Free Will a Cultural Universal?"
In: *Mind & Language* 25.3, pp. 346–358.
DOI: 10.1111/j.1468-0017.2010.01393.x
(Cit. on p. 5).
Savranskaya, Svetlana V. (2005).
"New Sources on the Role of Soviet Submarines in the Cuban Missile Crisis".
In: *Journal of Strategic Studies* 28.2, pp. 233–259.
DOI: 10.1080/01402390500088312
(Cit. on p. 19).
Schlick, Moritz (1939).
Problems of Ethics.
New York: Prentice Hall Inc.,
P. 243
(Cit. on p. 5).

Shimony, Abner (1993).
"Search for a worldview which can accommodate our knowledge of microphysics".
In: *The Search for a Naturalistic World View. Scientific method and epistemology*. Vol. 1.
New York: Cambridge University Press,
Pp. 62–76.
DOI: 10.1017/CBO9780511621147.003
(Cit. on p. 27).
Sober, Elliott (1999).
"Physicalism from a Probabilistic Point of View".
In: *Philosophical Studies* 95.1/2, pp. 135–174.
DOI: 10.1023/a:1004519608950
(Cit. on p. 34).
Speak, Daniel (2011).
"The Consequence Argument Revisited".
In: *The Oxford Handbook of Free Will*.
Ed. by Robert Kane.
2nd ed.
New York: Oxford University Press, Inc.
Chap. 6, pp. 115–130.
DOI: 10.1093/oxfordhb/9780195399691.003.0006
(Cit. on pp. 12, 17, 26).
Thompson, Valerie A. and Byrne, Ruth M. J. (2002).
"Reasoning counterfactually: Making inferences about things that didn't happen."
In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 28.6, pp. 1154–1170.
DOI: 10.1037/0278-7393.28.6.1154
(Cit. on p. 20).
Uffink, Jos (2007).
"Compendium of the foundations of classical statistical physics".
In: *Philosophy of Physics: Part B*.
Ed. by Jeremy Butterfield and John Earman.
Amsterdam: North Holland,
Pp. 923–1074
(Cit. on pp. 52, 57).
Vihvelin, Kadri (1998).
"John Martin Fischer, The Metaphysics of Free Will (Oxford: Blackwell, 1994)".
In: *Noûs* 32.3, pp. 406–420.
DOI: 10.1111/0029-4624.00107
(Cit. on p. 6).

Vihvelin, Kadri (2015).
"Arguments for Incompatibilism".
In: *The Stanford Encyclopedia of Philosophy*.
Ed. by Edward N. Zalta.
Fall 2015.
URL: http://plato.stanford.edu/archives/fall2015/entries/incompatibilism-arguments/ (visited on 10/10/2016)
(Cit. on pp. 4, 14).
Wallace, David (forthcoming).
"The Logic of the Past Hypothesis".
In: *Time's Arrow and the Probability Structure of the World*.
Ed. by B. Loewer, B. Weslake and E. Winsberg.
Cambridge, MA: Harvard University Press
(Cit. on pp. 50–52, 54, 55).
Wilson, Jessica (2014).
"Hume's Dictum and the Asymmetry of Counterfactual Dependence".
In: *Chance and Temporal Asymmetry*.
Ed. by Alistair Wilson.
Oxford University Press.
Chap. 13, pp. 258–279.
DOI: 10.1093/acprof:oso/9780199673421.003.0013
(Cit. on p. 45).
Wüthrich, Christian (2011).
"Can the World Be shown to be Indeterministic after all?"
In: *Probabilities in Physics*.
Oxford University Press,
Pp. 365–389.
DOI: 10.1093/acprof:oso/9780199577439.003.0014
(Cit. on pp. 28, 29).
Yoshimi, Jeffrey (2011).
"Supervenience, Dynamical Systems Theory, and Non-Reductive Physicalism".
In: *The British Journal for the Philosophy of Science* 63.2, pp. 373–398.
DOI: 10.1093/bjps/axr019
(Cit. on p. 34).



He also claims he can raise his hand whenever he wants! (De Mola 2016)

7 End Matter

7.1 Acknowledgments

Dear C., G., M., P., T., W.
the counterfactual

'Had you not supported me' $\Box\!\!\rightarrow$ 'this thesis would not have been written'

is almost trivially true. If you read this thesis you will learn why. But I am sure you know anyway, so thank you for everything.

I would also like to thank Prof. Landsman and Prof. Lüthy for their help during my time at Radboud and Prof. Slors for agreeing to be an examiner.

7.2 Certification of Ownership

Certification of ownership of the copyright in a typescript or manuscript

Thesis presented as part of, and in accordance with, the requirements for the Master of Arts Degree at Radboud University Nijmegen, Faculty of Philosophy, Theology and Religious Studies.

I hereby assert that I own exclusive copyright in the item named below. I give permission to the Radboud University Nijmegen Library to add this item to its stock and to make it available for consultation in the library, and for interlibrary lending for use in another library. It may be copied in full or in part for any bona fide library or research work, on the understanding that users are made aware of their obligations under copyright legislation, i.e. that no quotation and no information derived from it may be published without the author's prior consent.

Author	Davide De Mola
Title	Compatibilism and Actual Miracles
Date of submission	3rd December 2016

Signed (electronically):

Davide De Mola Southampton, 3rd December 2016

7.3 Acronyms

Det Determinism. 10

Essay An Essay on Free Will (van Inwagen 1983). 14

FAPP For All Practical Purposes (Quantum Mechanics), i.e. the recipe practising physicists use to model the outcomes of experiments.. 31

FWT Free Will Thesis. 9, 35

LMC Local Miracle Compatibilism. The variety of compatibilism defended by Lewis in reply to van Inwagen's Consequence argument. 13

M Claim linking determinism, counterfactuals and miracles. 20, 44

P_0 A true historical proposition. 15

PR1 Principle of Recombination of Possibilities on a reading friendly to van Inwagen (adapted to the Free Will Theorem scenario). 41

PR2 Principle of Recombination of Possibilities on a reading friendly to Lewis (adapted to the Free Will Theorem scenario). 42

S Ability to render a proposition false in the strong sense. 17

SDC (Wallace's) Simple Dynamical Conjecture (see equation 4.29f.). 51

SM Statistical Mechanics. 48

T3* Claim linking determinism, autonomy and sunflowers. 6, 50

UC Unavoidability Closure. 15

W Ability to render a proposition false in the weak sense. 17

***w*Sα** A world w is similar to the actual world α to degree d or more. 19

Faculty of Philosophy,
Theology and Religious Studies

Radboud University