Laws of Nature and Free Will

MERLUSSI, PEDRO

Laws of Nature and Free Will

Pedro Merlussi

A Thesis presented for the degree of
Doctor of Philosophy

Department of Philosophy
University of Durham
England

November 2017

Laws of nature and free will

Pedro Merlussi

Submitted for the degree of Doctor of Philosophy
November 2017

Abstract

This thesis investigates the conceptual relationship between laws of nature and free will. In order to clarify the discussion, I begin by distinguishing several questions with respect to the nature of a law: i) do the laws of nature cover everything that happens? ii) are they deterministic? iii) can there be exceptions to universal and deterministic laws? iv) do the laws of nature govern everything in the world? In order to answer these questions I look at three widely endorsed accounts of laws: "Humean" regularity accounts, laws as relations among universals, and the dispositional essentialist account. I argue that there is nothing in the very nature of a law - in any of the accounts surveyed - that implies a positive answer to questions (i) and (ii). I show that this has important consequences for the free will problem.

I then turn to the compatibility of free will and determinism. I focus on the Humean view and the dispositional essentialist account of laws. And the bulk of this discussion concerns the consequence argument, especially the question of whether the laws of nature are "up to us". I show that, on the dispositional conception of laws, there is no sense in which the laws of nature are up to us, contrary to the Humean view. However, this does not mean that there is no room for free will on the dispositional account. I argue that free will requires the laws of nature to be limited in scope, rather than being indeterministic. I conclude by showing that this allows one to resist the claim that indeterminism rules out free will.

Declaration

The work in this thesis is based on research carried out in the Department of Philosophy at the University of Durham between January 2014 and October 2017. No part of this thesis has been submitted elsewhere for any other degree or qualification and it is all my own work unless referenced to the contrary in the text.

Some parts of chapters 1 and 2 are based on joint research with Nancy Cartwright (Durham University/UCSD). The bulk of subsubsection "Laws as relations among universals: Extent and Permissiveness" (Chapter 2) was carried out by Nancy Cartwright.

Contents

Abstract ii

Declaration iii

1 Introduction 1

1.1 The problem of free will and determinism initially characterised . . . 1

1.1.1 Either determinism is true or not . . . 4

1.1.2 If determinism is true, there is no free will . . . 10

1.1.3 If determinism is not true, there is no free will . . . 12

1.1.4 There is free will . . . 16

1.2 An outline . . . 18

2 Are laws of nature consistent with contingency? 22

2.1 The "Humean" regularity account . . . 24

2.1.1 Extent . . . 26

2.1.2 Permissiveness . . . 27

2.1.3 Reliability . . . 28

2.1.4 Potency and Free Will . . . 32

2.2 Laws as relations among universals . . . 34

2.3 Dispositions, a la Alexander Bird . . . 40

2.4 Final remarks . . . 46

3 The no-choice transfer rules and the Lewis-Stalnaker semantics 48

3.1 The context of the discussion on rule (β) . . . 49

3.2 Widerker's counterexample to (β) . . . 53

3.3 Responding to the counterexample: strengthening rule (β) 57
3.4 McKay and Johnson's counterexample 60
3.5 Crisp and Warfield's defence of (β) 62
3.6 Against Crisp and Warfield 64
3.7 Rule (β) and the Lewis-Stalnaker semantics for counterfactuals 67
3.7.1 The counterfactual sufficiency interpretation 67
3.7.2 Lewis' theory and the *might*-counterfactual 68
3.7.3 **"M"** and (β) on Lewis' theory 71
3.7.4 A demonstration of $\mathbf{M}\phi$, $\mathbf{M}(\phi \supset \psi) \vdash \mathbf{M}\psi$ on Lewis' theory . 71
3.7.5 Stalnaker's theory and the limit assumption 75
3.7.6 Agglomeration on Stalnaker's theory 77
3.8 Final remarks 82

4 Local Miracle Compatibilism 83
4.1 Lewis' reply to van Inwagen's first formal argument 85
4.1.1 Counterfactuals and overall comparative similarity 88
4.1.2 Weak and strong abilities 91
4.1.3 Local Miracle Compatibilism 95
4.2 Beebee's objection to LMC 97
4.3 Interlude: Oakley's objection to Beebee 100
4.4 Is LMC compatible with Lewis' BSA? 106
4.4.1 The problem 106
4.5 LMC and contrastive explanation 108
4.6 Conclusion 118
4.7 Addendum: reply to Westphal 119
4.7.1 Westphal's objection 120

5 The dispositional conception of laws and agent causation 123
5.1 The charge of modal fallacy 124
5.2 Dispositional essentialism and laws 127
5.3 An objection to dispositional essentialism 132
5.4 Space-invading properties and Handfield's solution 134

5.4.1 An interlude: finks and antidotes . . . 135
5.4.2 Space-invaders . . . 139
5.5 Agent causation . . . 142
5.5.1 The metaphysics of agent causation . . . 143
5.5.2 Agent causation and space-invasion . . . 145
5.6 Final remarks . . . 149

6 Can we change the past? 151
6.1 Campbell's no past objection . . . 152
6.2 Time-indexing in the "**M**" operator . . . 156
6.3 NP and NN do not entail fatalism . . . 159
6.4 Arguing for premise 5 . . . 160
6.5 Bailey's objection to the necessity of the present . . . 161
6.5.1 Abstract times and Finch's trans-temporality thesis . . . 162
6.5.2 The necessity of origin and essentialist incompatibilism . . . 165
6.6 Final remarks . . . 167

7 The *Mind* (and luck) argument 168
7.1 Van Inwagen's formulation of the *Mind* Argument . . . 169
7.1.1 A libertarian solution: Finch and Warfield . . . 171
7.1.2 Nelkin: why the libertarian solution fails . . . 174
7.1.3 Graham: the discussion is confused . . . 175
7.2 Interlude: The luck argument . . . 182
7.3 Final remarks . . . 185

8 Conclusion 186

Appendix 190

A Better best system accounts 190
A.1 Van Fraassen's problem . . . 192
A.2 Recent attempts to solve van Fraassen's problem . . . 194
A.3 Cohen and Callender's account . . . 195

B Proofs of rules (α-M) and (β-M) 199

B.1 α-M 199

B.2 β-M 200

Bibliography 202

List of Figures

6.1 Circular time 155

Chapter 1

Introduction

Free will is, we all know, a huge question to which over millennia an enormous amount of intense thought has been dedicated, involving of course debate over the very formulation of the problem. The general topic to be discussed here is the conceptual relation between laws of nature and free will. This chapter is divided into two parts. First, I set forth the problem of free will and determinism. Second, I give an outline of the following chapters[1].

1.1 The problem of free will and determinism initially characterised

I have made a choice while typing these words. I have decided - perhaps from a diverse cluster of *possible* courses of action - to write this introduction. And I believe that this was up to me, that I had *control* over this decision. There are two main ideas here (Fischer 1999: 98). Free will, at least according to a certain traditional view, requires both (i) genuine alternative possibilities and (ii) control[2]. A nice way to picture this is to think of our future as *a garden of forking paths.*

[1] Parts of this chapter are based on a joint paper with Nancy Cartwright.

[2] Following John Fischer, I will use the term "control" because, as he says, "it highlights the fact that mere chance occurrences do not secure the satisfaction of the relevant requirement" (Fischer 1999: 99).

Many paths are genuinely available to us, and we can "select which path will be the path into the future" (Fischer 1999: 99).

Common sense suggests that there is free will. It is not totally clear, however, whether the very notion of free will is coherent. One worry comes from the thesis of determinism. If determinism is true, then it seems that there cannot be genuine alternative possibilities. On the other hand, philosophers worry that if determinism is not true, then our actions are "indeterministic" or "random", so that they cannot be under our control. It seems as though (i) and (ii) are in conflict. In order to have a better grasp of this, let us start by considering the notion of determinism. Since there are several theses that may be properly called "determinism", I shall start by distinguishing them.

First, one may reasonably assume that the principle of bivalence holds for all propositions, that is, that every proposition is either true or false. Now consider the proposition that I will have a beer tomorrow night. Either this proposition is true, or it is false. If it is true, then I will have a beer tomorrow night. If it is not true, then I will not have a beer tomorrow night. In any case, it seems that it is not up to me whether I will have a beer tomorrow night, since there is a true proposition describing what I am going to do in the future. This is what some philosophers call "determinism", or "logical determinism" (Jordan 1963; Lukasiewicz 1970). The problem of logical determinism is that of understanding how agents can have free will if there are true propositions about their future actions.

Second, the term "determinism" may be used to refer to the view according to which God knows every true proposition, including those about our future actions. In this sense, this view may be more accurately called "theological determinism". And it is called "determinism" because it may be the case that God's omniscience is incompatible with free will (Fischer & Tognazzini 2013). This is the problem of theological determinism and free will, that is, the problem of whether God's omniscience is compatible with free will.

Naturally, one may use the term "determinism" in a way that is neither "logical determinism" nor "theological determinism". It is the conception of determinism often called "nomological determinism" or "causal determinism". Very roughly,

nomological determinism is the thesis according to which the past and the laws of nature determine the future. Nomological determinism is also connected with free will, since it is arguably a threat to the view that there are genuine alternative possibilities. But it differs from the views mentioned above because it is deeply related to the concept of laws of nature and our comprehension of the physical sciences. It is this conception of determinism I shall be concerned with in this thesis. The main question I shall address here is: What is the relation between laws and free will?

Clarifying the problem. In order to answer these questions, I must explain what is the problem of free will and nomological determinism (henceforth just "determinism"). One of the ways to put forward a philosophical problem is in terms of a set of propositions all of which seem true (or for all of which we seem to have good reasons), but which are jointly inconsistent. We can sum up the problem of free will and determinism as follows:

1. Either determinism is true or not.

2. If determinism is true, then there is no free will.

3. If determinism is not true, then there is no free will either.

4. There is free will.

Propositions (2) and (3) are the controversial ones, but there are influential arguments in the literature for them. That is, as I will explain below, the consequence argument is an argument for (2), while the *Mind* argument and the luck argument are arguments for (3). If the consequence argument is cogent, then determinism rules out genuine alternative possibilities. On the other, if the *Mind* and the luck arguments are cogent, then the truth of indeterminism rules out control. So, if these arguments are cogent, there are good reasons for accepting (2) and (3). And there are also good reasons for propositions (1) and (4). (1) is an instance of a logical truth. With respect to (4), it seems no one can deny it, at least "until they started doing some philosophy" (Vihvelin 2015: 395).

Now the problem should be clear. For suppose determinism is true. If this is the case, given proposition (2), it follows that there is no free will. Now suppose that determinism is not true. If this is the case, given proposition (3), it follows that there is no free will. Therefore, in any case, there is no free will, which is the contradictory of proposition (4). So, as we can see, theses (1) to (3) entail the contradictory of (4). This is how the philosophical problem of free will and determinism may be initially presented.

As Kadri Vihvelin (2015) points out, the fact that we have free will seems undeniable until someone starts doing philosophy. Naturally, the philosophical problem may be presented as a dilemma against (4). Either determinism is true or not. If determinism is true, then there is no free will. If determinism is not true, then there is no free will either. Thus, there is no free will.

All of this requires a more precise characterisation. I have not sufficiently explained what I mean by "determinism", nor have I explicitly presented the arguments for (2) and (3). On top of that, there are some important questions with respect to the notion of determinism that need to be distinguished. I shall explain the dilemma in a bit more detail in the following sections, as well as the arguments for (2) and (3).

1.1.1 Either determinism is true or not

Determinism, as I mentioned before, is deeply related to the concept of laws of nature. But it is not my aim to provide an original account of the sort of thing a law of nature is. Rather, I shall look at some widely endorsed accounts in chapter 2 in order to discuss their connection with free will. There is, however, a way of cashing out the notion of "determinism" that one may take as independent of any particular philosophical account of lawhood. Call it "the standard view". It may be presented as follows (van Inwagen 1983: 65; Vihvelin 2013):

> Determinism is the thesis that for every instant of time t, there is a proposition that expresses the state of the world at that instant, and if P and Q are any propositions that express the state of the world at some instants, then the conjunction of P together with the laws of nature entails Q.

This definition allows us to elucidate the intuitive idea about nomological determinism, namely, that the past and the laws determine the future. If P expresses the state of the world at t and Q expresses the state of the world at some time t' later than t, then one part of the world's history (namely, the past) together with the laws determine another part of that history (namely, the future). This definition has also the consequence that the future and the laws determine the past, for P may be a proposition that expresses the state of the world at an instant t later than the one expressed by Q.

This is a legitimate way of cashing out the notion of determinism. But there are two theses in this formulation that need to be properly distinguished.

Extent and permissiveness

Vihvelin does a nice job pointing out that the standard sense of determinism, roughly stated, is the conjunction of the following two claims: (i) that the laws hold everywhere and everywhen, covering all aspects of what happens[3]; (ii) "that the laws state *sufficient*, as opposed to merely necessary or probabilistic, conditions" (Vihvelin 2013: 3). Following work with Nancy Cartwright, I shall keep the following two questions separate from one another:

- *Extent.* Is everything that happens covered by the laws of nature? For instance, there may be happenings, or kinds of happenings, or whole domains about which L - the complete set of correct laws - is silent.

- *Permissiveness.* When L speaks about the outcomes that are to occur, what kind of latitude does it admit? For instance, does it always select a single happening? Does it always lay down at least a probability, or can L admit a set of different outcomes, remaining silent about their probabilities[4]?

[3] Actually, Vihvelin formulates (i) like this: "(i) that we are *no exception* to the laws that govern everything else in the universe". I avoided this formulation because the word "govern" is a bit misleading, even though Vihvelin herself is aware of this.

[4] Cartwright and I propose treating the laws that say "anything goes" in some circumstance as not covering that circumstance and thus limited in extent.

To explain these questions I must first say something about laws. We may very broadly think of laws as saying something about the occurrence or features of some kinds of events given other facts about those events or others. I shall call events of the former kind, the "domain of the law", those of the latter, the "input". The free will discussion focuses on laws where the input for any event described by P in the law's domain is in P's past. I will designate an input for P that is allowed under law L and that occurs at t_0 by $P_{t0}(L, P)$.

There are two guiding ideas I rely on throughout in considering *extent* and *permissiveness*. The idea for *extent* is simple. There may be situations where the laws are silent; they simply do not cover those situations. This is an issue that is orthogonal to questions about whether laws are permissive when they do speak. For instance, L may be deterministic in the sense that for each $P_{t0}(L, P)$, L admits one and only one P to occur, yet limited in extent because some real situations are not $P_{t0}(L, P)$-type situations for any admissible $P_{t0}(L, P)$, i.e. some situations may not fall into any of the categories for the additional facts that bring L to bear.

With respect to extent, a little simple housekeeping is necessary since some of the discussion in both the philosophy of science and the free will literature as well as in the related metaphysics literature is confusing (at least to Cartwright and me) because it does not make clear the formulations at stake to begin with, especially with respect to the quantifiers and what they range over. Consider the claim G': "Politeness requires giving an expensive gift to one's teacher/mentor", that I suppose is true in some cultures influenced by Confucianism. Shall we say it is limited in extent, or shall we rather consider G: "In cultures A,B,C, politeness requires giving an expensive gift to one's teacher/mentor", which is, I suppose, true everywhere and in that sense not limited in extent? Similarly Cartwright (1983, 1989, 2009, 2010) has suggested that what we think of as the usual laws of physics L' may well be limited in extent in a very specific way: They may be unable to represent all the possible causes of the effects they represent; their truth may then be restricted to just those cases where only causes they can represent are at work. Thus, it should more perspicaciously be formulated something like this. L: "So long as all of the causes of the consequences represented in L' are features represented in the

antecedents of L', then L'". One could think of formulating the issue in terms of domain restrictions: Are these restrictions included in the laws themselves or not? The problem is that it can be difficult to formulate criteria for what counts as a restriction on the domain of a law versus what counts as a feature that it genuinely covers. This is why I formulate the issue as follows: Are there things that happen that the complete and correct set of laws does not cover?

As to *permissiveness*, Cartwright has long urged that some events may just happen - by hap - without even any probabilities assigned by nature. An ear-ring back is stuck in some debris in the crack between the floorboards. Someone tries to lift it with a magnet. The magnet pulls upward on the metal object with a fixed strength and gravity pulls it down with a fixed strength. These activities are both properly treated as sources of forces, where by "properly" she means that there is a general way to ascribe forces for both. There is a magnet and there is a rule in physics for what forces magnets exert; and there is a large mass – the earth – and there is a rule for what force a mass exerts. There is also debris that inhibits the motion of the ear-ring back. Maybe there is another description of this particular debris for which there is a proper rule in physics that assigns a force. But certainly not under the description "debris". And maybe there is no other such description. We may grant that some causes of motion are forces in the proper sense of that concept but that does not imply that all are. To assume there must be because the debris can affect the motion of the ear-ring back is to make a massive metaphysical assumption beyond the empirical evidence, Cartwright (2000, 2010) argues (cf. Cartwright & Pemberton 2013).

If one leaves the issue open, then a new possibility arises. There is a rule for what force is exerted when the magnet and the earth act together in this arrangement, and on this rule only one resultant force is allowed. But what about the motion of the ear-ring back? Is there a rule that says what one motion will happen in this arrangement when the resultant force of the earth and the magnet acts on the ear-ring back simultaneously with the inhibiting power of the debris, or if not a rule dictating one single outcome, is there a rule that dictates a set of outcomes with a probability measure over them?

It seems we have insufficient reason to assume there is, Cartwright has argued, so that assumption should not be forced by our account of what laws are; the account should leave the question open. Yet surely there is some kind of rule since we have what do seem well-warranted beliefs that the ear-ring back will not fly away at near the speed of light. This is very underexplored territory. Cartwright and I have labelled this *permissiveness*: When L applies, given a relevant input $P_{t0}(L, P)$, L might admit only one outcome, in which case L is not permissive. On the other hand, L may be permissive in the sense that L admits a set that includes more than one outcome, and in the latter case, L may or may not provide a probability over that set[5].

What we want then is a definition of "deterministic" that is sensitive to these aspects. A definition in which - for instance - laws may be deterministic and yet limited in *extent*[6]. Now, there are different ways to characterise "deterministic" (see, for example, Müller and Placek 2015). Here, I shall discuss what they call the "mapping-based approach", which springs from Richard Montague's (1974) semantic characterisation of deterministic theories. I follow, in particular, John Earman's work (1986), which strips away "much of the formal apparatus employed by Montague" (Roberts 2006: 199) in characterising determinism in terms of physically possible worlds. That is, the strategy is to define "deterministic" by quantifying over all the possible worlds allowed by the set of laws.

Let $\mathcal{L}$ to stand for "L is the correct set of laws". Here is how "deterministic" is defined:

Definition 1.1.1. *Laws L are deterministic if, and only if, for any P that L covers and any $P_{t0}(L, P)$ that is an input to L for P and any logically possible worlds w, w' in which $\mathcal{L}$, if w and w' agree on $P_{t0}(L, P)$, they agree on whether P obtains.*

[5] Clearly this supposes some already given way to individuate outcomes.

[6] Another motivation for keeping the questions about extent and permissiveness separate from one another is that the standard view is at odds with some recent accounts of laws. Two of them are Jonathan Cohen and Craig Callender's (2009) "Better Best Systems Account" and Barry Loewer's view (2007). I show in the appendix that the laws of nature may not hold everywhere and everywhen, and yet there is a sense in which one might say that they are deterministic.

This definition allows us to keep the questions about *extent* and *permissiveness* separate from one another. According to the standard view, the claim that determinism is true boils down to saying that laws have universal *extent* and are *deterministic*.

Perhaps one might question the relevance of such a distinction because in the relevant sense of "laws" - such as David Lewis' best system account (1973, 1994) or David Armstrong's necessitarian view (1983, 1997) - they are universal in *extent*. However, I shall argue in chapter 2 that this is not the case even for the mainstream accounts: the claim that laws are universal in *extent* is just an add-on to these accounts of laws.

This has some consequences for my formulation of the compatibility problem of free will and deterministic laws. For strictly speaking the problem with respect to the laws of nature and free will is not about free will and deterministic laws, at least not according to the definition I employ. If it turns out that laws are limited in *extent* and do not apply to actional-events, then they do not seem to be a threat to free will. The claim that laws are a threat to free will depends on the assumption that they are universal in *extent*, or at least that they cover actional-events. In this sense, the first proposition of the dilemma should be more explicitly formulated as follows: either laws are deterministic *and* universal, or not. So in a nutshell, the problem is about whether deterministic laws *together with* the claim that laws are not limited in *extent* (and thus cover actional-events) is a threat to free will.

Furthermore, distinguishing these questions opens the possibility of differentiating at least two views with respect to the compatibility of free will and indeterminism - where indeterminism is the denial of determinism. For example, it is usually argued that if determinism is incompatible with free will, then if there is free will, the laws of nature have to be indeterministic. (I suspect that this is so because a good part of this discussion seems to presuppose that the laws of nature cover, or perhaps govern, everything in the world). Nevertheless, if we distinguish the questions about *extent* and *permissiveness*, then if free will and determinism are incompatible and there is free will, what follows is that *either* laws are indeterministic *or* they are limited in extent. Someone may be an incompatibilist in the sense that deterministic

and universal laws are incompatible with free will, but deny that free will requires laws to be indeterministic. I will articulate such a view in chapter 5.

1.1.2 If determinism is true, there is no free will

There are many arguments for the incompatibility of free will and deterministic laws insofar as determinism is the claim that laws are universal in *extent* and deterministic[7]. I will characterise incompatibilism as the thesis that, necessarily, if laws are deterministic and universal in extent, there is no free will. Let *compatibilism* be the denial of incompatibilism; that is, it is possible for laws to be deterministic, universal in extent, and nevertheless there is free will. Among the arguments for incompatibilism there is no doubt that the consequence argument - defended most famously by Peter van Inwagen (1983) and Carl Ginet (1983) - is the most influential one. As Neil Levy and Michael McKenna write "the consequence argument is very powerful. It has been credited with breaking the compatibilist hegemony over the free will debate" (Levy & McKenna 2009: 97). The informal formulation of the argument runs as follows:

> If determinism is true, then our acts are the consequence of laws of nature and events in the remote past. But it's not up to us what went on before we were born, and neither is it up to us what the laws of nature are. Therefore, the consequences of these things (including our present acts) are not up to us. (van Inwagen 1983: 16)

There are many attempts to fill in the details of this enthymeme, and I will discuss some of them in the following chapters[8]. The main idea is simple. Let us assume that our acts are covered by the correct set of laws. Now consider the following question:

- *Reliability.* Does what L (plus some relevant $P_{t0}(L, P)$) says is to happen always happen? For instance, can there be exceptions to L and yet L still be the correct and complete set of laws?

[7] See Vihvelin (2013: chapter 5) for a nice survey of the traditional arguments for incompatibilism. Here I shall be mainly concerned with the consequence argument.

[8] In particular, van Inwagen's first and third formal arguments, and Alicia Finch and Ted Warfield's (1998) formulation.

Suppose P is a proposition such as the one expressed by the sentence "I raise my hand". Given that L - the correct set of laws covers P, if $\mathcal{L}$ and L is deterministic, then laws are *reliable*: what L says to happen always happen. This can be formulated in a more precise way as follows:

D: If L is deterministic, $\Box((\mathcal{L} \wedge P_{t0}(L, P)) \supset P)$ follows.

To see why, suppose that L is deterministic and $P_{t0}(L, P)$ is an initial condition for P with respect to L and P. Let $\mathcal{W}$ stand for the collection of all possible worlds. Consider an arbitrary world w in $\mathcal{W}$ where $P_{t0}(L, P)$ and $\mathcal{L}$. Because L is deterministic, if P obtains in any world where $\mathcal{L}$ and $P_{t0}(L, P)$ obtain, it holds in all worlds where $P_{t0}(L, P)$ and $\mathcal{L}$ obtain, including w. P obtains in our (the actual) world where L and $P_{t0}(L, P)$ obtain. So P obtains in w and thus $(P_{t0}(L, P) \wedge \mathcal{L}) \supset P$ in w. Since w is any arbitrary possible world, $\Box((\mathcal{L} \wedge P_0(L, P)) \supset P)$ follows.

Now, the idea of the consequence argument is that neither the fact that $\mathcal{L}$ holds nor the fact that $P_{t0}(L, P)$ holds are *up to us*. Both Ginet and van Inwagen have introduced a modal sentential operator in order to cash that out. The operator is "**N**", and (according to van Inwagen's formulation) $\mathbf{N}\phi$ stands for "ϕ and no one has or ever had any choice about whether ϕ". Furthermore, "**N**" is supposed to satisfy the following inference rules:

(α) $\Box\phi \vdash \mathbf{N}\phi$

(β) $\mathbf{N}\phi$, $\mathbf{N}(\phi \supset \psi) \vdash \mathbf{N}\psi$

Here is the consequence argument for incompatibilism. If the the laws of nature are deterministic and universal, they cover P (say, that I raise my hand).

1	$\Box((\mathcal{L} \wedge P_{t0}(L, P)) \supset P)$	def 1.1.1
2	$\Box(\mathcal{L} \supset (P_{t0}(L, P) \supset P))$	modal logic, 1
3	$\mathbf{N}(\mathcal{L} \supset (P_{t0}(L, P) \supset P))$	α, 2
4	$\mathbf{N}\mathcal{L}$	premise
5	$\mathbf{N}P_{t0}(L, P)$	premise
6	$\mathbf{N}(P_{t0}(L, P) \supset P)$	β, 3, 4
7	$\mathbf{N}P$	β, 5, 6

The consequence argument is an argument for the view that if laws are deterministic and universal in extent, then no one is able to do otherwise[9]. Remember that, according to the traditional view, free will requires genuine alternative possibilities. So, if the consequence argument is cogent, there cannot be genuine alternative possibilities, so that determinism is incompatible with free will[10]. And if this is the case, the laws of nature cannot be both deterministic and universal if we believe some agents have free will.

1.1.3 If determinism is not true, there is no free will

Suppose incompatibilism is true. If that is the case, is there any room for free will? Call "libertarianism" the view that incompatibilism is true and there is free will. If libertarianism is right, then indeterminism (which is the denial of determinism) is a necessary condition for free will. However, it is worth pointing out that even libertarians do not take for granted that indeterminism *per se* guarantees the existence

[9] Well, at least this is what I will show by defining "**N**" in order to demonstrate the validity of the argument. And I guess that this definition (roughly) boils down to "no one is able to do otherwise".

[10] This formulation, however, faces lots of problems. First, there is the problem of whether (β) is a valid inference rule. Second, since incompatibilism is the thesis that, *necessarily*, if determinism is true (in the sense that laws are deterministic and universal in extent), then there is no free will, the argument cannot rely on premises that are merely actually true, such as premises 4 and 5 (Warfield 2000). In any case, however, I believe that there is *some* formulation of the argument which is immune to these critiques, as I shall argue in chapters 3, 5 and 6.

of free will. For instance, if a long time ago some particles in a far, far away galaxy were undetermined and everything else was determined, there would not be such a thing as free will (assuming, of course, that determinism and free will are incompatible). In other words, if the laws for non-agents are permissive, but impermissive as long as agents are involved, then there is no free will (simply because there are no genuine alternative possibilities). What is important for those holding that our having free will requires indeterminism is that either the laws of nature do not cover our actions, or if they cover our actions, they are indeterministic.

However, the term "indeterminism" seems to have something to do with luck and randomness; and luck and randomness do seem incompatible with free will because they seem incompatible with *control*. And if this is the case, then free will is incompatible with indeterminism after all. The *Mind* and luck arguments are attempts to turn this intuition about indeterminism and free will into precise rigorous arguments.

The *Mind* argument

The *Mind* argument has received this name because van Inwagen noticed that it has appeared very often in the philosophy journal *Mind*[11]. Here, I will focus on van Inwagen's version of the argument (in particular the third strand) and its subsequent discussion (Finch & Warfield 1998; Nelkin 2001; Graham 2014; Shabo 2013). It is controversial, however, whether the very formulation of the argument is not muddled (Graham 2010). I will follow the standard way of putting forward the argument anyway, leaving the question of whether this discussion is confused to chapter 7.

In *An Essay on Free will* van Inwagen sums up the argument as follows: "a free act is an act one has a choice about; but no one has any choice about that which is undetermined" (van Inwagen 1983: 142). But why exactly no one has any choice about that which is undetermined? He defends this claim by asking us to imagine a mechanism the salient features of which are a red light, a green light, and

[11]It appeared for example in Hobart (1934), Nowell-Smith (1948, 1957), Smart (1961), Ayer (1954), among others.

a button. If someone presses the button, the mechanism will result in one of the lights flashing, but it is *undetermined* which. Following Christopher Franklin (2011: 225), I will assume that by "undetermined" van Inwagen means that the laws are indeterministic with respect to the light to which the light will flash. "Now suppose that you must press the button on this mechanism. Have you any choice about which of the lights will flash? It seems obvious that you have no choice about this" (van Inwagen 1983: 142).

Now consider the following scenario. Suppose that Mary is undecided about whether or not she should take a break from writing her PhD thesis. If laws are permissive with respect to this, she may decide to take a break and procrastinate a little bit, or she may decide to stay in her office working for a few hours more, etc. Suppose she thinks the matter over carefully and decides to keep on writing. Why did she do this? In order to finish her work, go back home early and chat with her friends.

In that scenario there is an explanation as to why Mary decided to stay in the office working; the explanation reveals Mary's reasons for performing that action. But what makes such an explanation true? According to *causalism*, for instance, there must be a causal connection between the *explanans* (say, the agent's desires, beliefs, etc.) and the *explanandum* (the agent's action). That is, the agent's reasons (Mary's desire to go back home early and chat with her friends) need to cause the action (the decision to stay in the office). Suppose the following is right, that Mary's reasons caused her action. Now, let P stand for the proposition expressed by the sentence "Mary decides to stay in the office" and let DB stand for the relevant desires and beliefs that do the causing[12].

The cogency of the consequence argument implies that there are no genuine alternative possibilities. But suppose that, in order to allow genuine alternative possibilities, we require laws to be indeterministic with respect to our actions. Let

[12]The argument can be formulated without presupposing *causalism*. But since in the original formulation van Inwagen (1983: 145) adopts a Davidsonian account sketched in "Actions, Reasons and Causes" (Davidson 1963), and since Davidson himself defended causalism in that paper, I think this formulation is fine for the present purposes.

us say, for example, that the laws of nature are indeterministic with respect to P being followed by DB, so that it is undetermined whether P is followed by DB. Here is a version of the argument (Finch & Warfield 1998):

1. $\mathbf{N}(DB \supset P)$
2. $\mathbf{N}DB$

   Thus,

3. $\mathbf{N}P$ (from 1, 2, and (β))

Now consider premise 1. Mary seems to have no choice about whether P follows DB in the same way that no one has any choice about whether the device will flash, say, a green light (for the laws of nature are indeterministic with respect to this). And it is also assumed that premise 2 is plausible[13]. If one is an incompatibilist because she is convinced that the consequence argument (as presented before) is cogent, then one will have trouble with the *Mind* argument. For if the consequence argument is cogent, rule (β) is valid, and in that case the *Mind* argument is valid as well. In fact, the purpose of bringing this formulation of the *Mind* argument into the discussion is that it relies on the assumption that "having no-choice about" transfers across material conditionals. As I shall argue in chapter 7, however, this discussion is not very clear and at least *this* formulation of the argument is not compelling.

The luck argument

The luck argument has recently received a lot of attention in the literature on free will. There is no single argument that is "the luck argument". There are rather a great many arguments *about* the connection between luck and free will. Franklin

[13] As Franklin says, this assumption is harmless because Mary "could only have had a choice about DB [...] if [s]he performed some earlier action which itself would have been brought about by yet earlier mental states. We could then raise the same questions about this earlier action and these earlier mental states. Someone might again insist that the agent had a choice about these still earlier metal states. But this cannot go on forever and we will eventually discover [Mary's] 'initial' mental states for which [s]he had no choice" (Franklin 2011: footnote 35).

(2011) identifies several formulations, and here I will be concerned with Alfred Mele's (2006) explanatory formulation.

In his original formulation, Mele conceives a scenario in which a goddess, Diana, creates "agents in an indeterministic universe which whenever they freely perform an action of deciding to a, they could have freely performed some alternative action" (Mele 2006: 8). Notwithstanding,

> She worries that her design does not accommodate this. Her worry, more specifically, is that if the difference between the actual world, in which one of her agents judges it best to A straightaway and then, at t, decides accordingly, and any possible world with the same past up to t and the same laws of nature in which he makes an alternative decision while the judgment persists is just a matter of luck, then he does not freely make that decision in that possible world, W. Diana suspects that his making that alternative decision rather than deciding in accordance with his best judgment— that is, that difference between W and the actual world—is just a matter of bad luck or, more precisely, of worse luck in W for the agent than in the actual world. After all, because the worlds do not diverge before the agent decides, there is no difference in them to account for the difference in decisions (Mele 2006: 8).

This what I will refer to as "the luck argument", which is supposed to be a problem for the compatibility of the denial of determinism and free will.

1.1.4 There is free will

So far I have motivated the most controversial assumptions of the problem (although I have not presented the arguments in complete detail yet). The claim that there is free will should be the least controversial one. One reason for believing in free will is that one may think it is connected to moral responsibility. If there is no free will, on many standard understandings, there is no moral responsibility. And because there is moral responsibility it follows there is free will.

Now, I do not know how to give a consensual characterisation of "free will", let alone a widely accepted *definition* of this term. Here, I will stick with the characterisation I started at the beginning. Fischer calls it "alternative-possibilities control":

> The intuitive picture behind the alternative-possibilities control requirement is that moral responsibility requires that the agent select one from among various genuinely open paths the world might take. There are two important ideas

> here. One is that there must be various paths genuinely available to the agent (at least at some times suitably related to the time of the behaviour under consideration). The second idea is that the agent (and not some outside force or mere chance) selects which path will be the path into the future. It seems to me that both ideas are important components of the traditional conception of the sort of control associated with moral responsibility - alternative-possibilities control. (Fischer 1999: 99)

I have avoided to use the term "moral responsibility" because there is a side of the debate I am not considering. If this intuitive picture is correct, we have the following:

Alternative possibilities: If no one is ever able to do otherwise, then there is no free will.

Moral responsibility If there is no free will, no one is morally responsible.

This is hugely contentious. If the conditionals presented above are true, one can infer the following:

PAP: If no one is ever able to do otherwise, then no one is morally responsible.

The assumptions are in line with what Carolina Sartorio[14] calls "the traditional view of freedom" (Sartorio forthcoming). In "Alternate Possibilities and Moral Responsibility" Harry Frankfurt (1969, 2003) has decisively influenced the debate by arguing against (PAP). Vihvelin compares the impact of Frankfurt's argument to Edmund Gettier's counterexamples to knowledge as justified true belief (Vihvelin 2013: 93). The scenarios Frankfurt appealed to have been called "Frankfurt-style examples". The difference, however, is that it is *still* controversial whether Frankfurt provided a successful refutation of (PAP).

As I said, I am not considering this side of the debate. Precisely for this reason I am considering the discussion concerning the "consequence argument", and not the "direct argument"[15]. If it turns out Frankfurt is right, then my formulation of

[14] See Sartorio's paper (forthcoming) on Frankfurt-style counterexamples for a nice survey.

[15] The direct argument, to use Levy and McKenna's expression (2009), is the "cousin" of the consequence argument. While the consequence argument relies on transfer principles with respect to "no-choice about", the direct argument works with the notion of *no-responsibility for*.

the consequence argument will not expose the incompatibility of determinism and moral responsibility, but just the incompatibility of determinism and free will, where "free will" - as I understand it - requires both genuine alternative possibilities and control. So, although the thesis is about laws of nature and free will, it is not about laws of nature and moral responsibility[16].

1.2 An outline

I have specified the general problem to be discussed throughout the thesis. But I have said too little about what laws are. In chapter 2 I look at some widely endorsed philosophical accounts of laws of nature: "Humean" regularity accounts, laws as relations among universals, and disposition/powers account. Chapter 2 has two purposes: to distinguish several important questions in formulating the problem of free will and to discuss how the term "deterministic laws" should be defined relative to the philosophical accounts discussed. My claim will be that "Humeans" and dispositional essentialist advocates should adopt different definitions of "deterministic laws". I motivate the definitions by discussing Scott Sehon's (2011) objection to the claim that deterministic laws are reliable.

The purpose of presenting both the consequence and *Mind* arguments in this chapter was to highlight the role "**N**" and rule (β) play in the debate, since the

[16]In the *Stanford Encyclopedia*'s entry for arguments for incompatibilism, Vihvelin distinguishes questions about free will from questions about moral responsibility. According to her, "Someone might believe that we have free will and that free will is compatible with determinism while also believing, for other reasons, that no one is ever morally responsible. And someone might believe that we don't have free will (because of determinism or something else) while also believing, against conventional wisdom, that we are nevertheless morally responsible. What one believes about determinism and moral responsibility will depend, in large part, on what one believes about various matters within the scope of ethics rather than metaphysics. Among other things, it will depend on what one takes moral responsibility to be (P. Strawson 1962; G. Strawson 1986, 1994; Watson 2004). For these reasons it is important not to conflate the question of the compatibility of free will and determinism with the question of whether moral responsibility is compatible with determinism." (Vihvelin 2017)

formulations of the arguments depend on this inference rule: $\mathbf{N}(\phi \supset \psi)$, $\mathbf{N}\phi \vdash \mathbf{N}\psi$. In chapter 3 I present the back and forth of the discussion concerning the validity of rule (β). I will discuss the main counterexamples against (β), as well as Crisp and Warfield's solution (2000) to these counterexamples. I put forward an argument from analogy to show that Crisp and Warfield's strategy is implausible. After that, I give two interpretations of "having no choice about whether a proposition is true" and prove that some "no-choice about" transfer rules (such as β) hold on the Lewis-Stalnaker semantics for counterfactuals. One interesting result is that, according to one of the definitions, (β) is valid on Stalnaker's semantics but not on Lewis'. After defining "deterministic laws" and finding some interpretation of "no-choice about" in which the inference rules hold, I go on to discuss the conceptual relationship between universal and deterministic laws and free will according to two accounts of laws described in Chapter 2, namely, the Humean view and Alexander Bird's dispositional view. The reason is that I take the Humean position on laws to be the best way to motivate compatibilism, and the dispositionalist conception of laws to be the best way to motivate incompatibilism.

In Chapter 4 I argue that at least some Humean compatibilist view motivates the rejection of premise 2 in the dilemma against free will. I discuss David Lewis' Local Miracle Compatibilism (LMC). Helen Beebee has interestingly argued that LMC is an untenable view because it depends on a quite problematic distinction. However, I will argue the distinction makes sense if the principle that a freely performed action requires a contrastive explanation (an explanation of why the agent performed a rather than not-a) is true.

In chapter 5 I argue that the dispositional essentialist account of laws and incompatibilism go hand in hand. I show that the dispositional account of laws entails the necessary truth of the premise that the laws of nature are not up to anyone. Nevertheless, my formulation of the consequence argument depends on the Lewis-Stalnaker semantics for counterfactuals, which seems incompatible with dispositional essentialism. I respond to this problem by appealing to Toby Handfield's solution (2001, 2004) according to which dispositional essentialists may appeal to the concept of a space-invading property instance. After that, I introduce the theory of agent

causation and draw some parallels between space-invasion and agent causation. And if agent causation can be conceived as space-invasion, then laws - according to the dispositional view - will have to be limited in extent rather than indeterministic if they are consistent with free will.

If the arguments in chapter 5 are sound, then incompatibilists should take dispositional essentialism seriously. But there is another premise in the consequence argument that requires justification, namely, the premise that no one can change the past. In chapter 6 I argue that the main counterexample to this premise - presented by Joseph Campbell (2007, 2008, 2010) - fails.

Finally, if the dispositional account of laws is correct and the premise that no one can change the past is true, free will is incompatible with laws being deterministic and universal in *extent*. The challenge then is to reply to the *Mind* argument. I argue that the standard formulation of the *Mind* argument fails. After that, I briefly consider the implications of the view articulated in chapter 5 with respect to the luck argument. I will argue that there is a gap in the argument, and that presupposes that the laws are both universal and indeterministic.

The purpose of bringing the Humean account of laws and the dispositional view of laws into play was to point out that different philosophical views with respect to laws have different implications for the question of whether determinism and free will are compatible. But if my overall argument is sound, at least one proposition of the dilemma against free will is unjustified. I do not claim, however, that I have a solution to the troubling problem of free will. A solution to the problem - I suggest - depends on explaining what is to have the degree of control necessary for free will, which is something I will not consider in detail here. In any case, I hope at least to clarify that there is nothing in the very nature of a law - in any of the accounts I survey - that implies that there are no genuine alternative possibilities. The only difference is that the Humean view is compatible with alternative possibilities even if determinism is true, while the dispositional essentialist view requires determinism not to be true. However, showing that laws are consistent with various paths available to the agent does not show that the agent has *control*, that is, that "the agent (and not some outside force or mere chance) selects which path will be the path into

the future" (Fischer 1999: 99). The solution to this problem - I suggest - does not depend on the question of whether determinism is true. It depends on what control consists in.

Chapter 2

Are laws of nature consistent with contingency?

In the previous chapter I introduced three questions concerning the nature of laws, namely, questions about *extent*, *permissiveness* and *reliability*. Cartwright and I have labelled the possibility of laws being *permissive*, limited in *extent* or *unreliable* as sources of *contingency* about what happens in the world[1]. My claim is that the questions matter because the very formulation of the free will problem relies on them; for instance, incompatibilism has bite as long as laws do not allow for contingency in the aforementioned senses. So, the general question to be dealt with in this chapter is that of whether laws are consistent with contingency. That is, does the existence of laws imply that things could not happen other than the way they do consistent with the laws staying the same?

The answer to the question of whether laws allow for contingency depends on what the laws of nature actually are, but it also depends on what they are like. This latter is my concern here. Different philosophical views give different accounts of the sort of thing a law of nature is. I shall look at three that are widely endorsed: "Humean" regularity accounts, laws as relations among universals, and the disposi-

[1]Parts of this chapter have been published as "Determinism, Laws of Nature and the Consequence Argument" (2016), and some other parts will be published as "Are laws of nature consistent with contingency?", which is a joint paper with Nancy Cartwright. The bulk of subsubsection "Laws as relations among universals: Extent and Permissiveness" was carried out by Cartwright.

tional essentialist account. The question is, given an account of what laws are, what follows about how much contingency, and of what kinds, laws allow?

Naturally, whether contingency is possible given the laws of nature depends not only on what kinds of things laws of nature are but also on what contingency consists in. I will introduce two more questions in addition to those I talked about in the previous chapter. Here are the questions to keep sorted from one another.

- *Extent.* Is everything that happens covered by L? For instance, there may be happenings, or kinds of happenings, or whole domains about which L is silent.

- *Permissiveness.* When L speaks about the outcomes that are to occur, what kind of latitude does it admit? For instance, does it always select a single happening? Does it always lay down at least a probability, or can L admit a set of different outcomes, remaining silent about their probabilities?

- *Reliability.* Does what L (plus some relevant $P_{t0}(L, P)$) says is to happen always happen? For instance, can there be exceptions to L and yet L still be the correct and complete set of laws?

- *Potency.* Do the things that L speaks about happen on account of L? Or, for instance, merely in accord with L?

- *Free will.* If P is an action of a person, is $\sim P$ consistent with $P_{t0}(L, P)$ and L being the correct and complete set of laws?

The last question - as one might expect - hovers in the background. One might take laws to be a threat to free will because it is assumed that they are deterministic. However, this should be reformulated as follows. Suppose that laws do not allow for contingency in the sense that (i) there are no happenings about which laws are silent, (ii) they are impermissive for agents and (iii) what they say to happen (with some relevant input) always happens. If (i), (ii) and (iii) hold water and the consequence argument is sound, then laws do not allow for contingency in the *free will* sense. Without these assumptions, though, it is hard to see how incompatibilism has bite. All these assumptions, however, can be envisaged as add-ons of the accounts I look at, or so I shall argue.

I will not always have much to say about every question with respect to each view of laws I survey but rather focus on what might not be altogether obvious or on where interesting differences lie. I will not focus too much on potency - though I list it for completeness and to make clear that it is a separate issue from the others. It is generally supposed – though not without objections – that universals and powers accounts allow for potency, as well as accounts that involve "necessary" regularities, whereas "Humean" regularity accounts do not.

The chapter is structured as follows. In section 2.1 I look at the best system account of laws (BSA for short), and, in particular, David Lewis' BSA. (In the appendix, I highlight some of the problems with it and examine two recent attempts to overcome it). Sections 2.2 and 2.3 concern - respectively - David Armstrong's account and the dispostionalist view of laws. The investigation here suggests that the dispositionalist conception of laws may rule out contingency but it need not. Conversely, the other accounts may admit contingency but they need not. In all three cases, I shall argue, the root idea of what laws are does not settle the issue of whether they allow contingency. Advocates of the different accounts may argue for one view or another on the issue, but (at least as I understand the accounts) this will be an *add-on* rather than a consequence of the basic view about what laws are.

2.1 The "Humean" regularity account

The central motivating idea behind what I shall call the "Humean" regularity account of laws is not about laws but about the make-up of the world. The facts that constitute the world involve only qualities, quantities, and relations that are occurent, where "occurent" means different things to different philosophers who call themselves "Humeans". What they all have in common is that they want to exclude any kind of modal features. There are no causings, no necessitatings, no doings, no making-things-happen-ings.

In answer to the question "What is it to be a law of nature?" the naive "Humean" account states that laws are regular associations among occurent features. But this is thought to be problematic. There are true accidental regularities that are not

laws, it is supposed. To use Hans Reichenbach's memorable example (1947: 368), "All gold spheres are less than a mile in diameter" is a genuine regular association, but this does not seem to be a law. So, it is commonly assumed, a satisfactory "Humean" view of laws should distinguish laws from accidental regularities. This is what David Lewis's best system account (BSA) sets out to do (Lewis 1973). Since the BSA is very well developed and widely adopted, I shall focus on this version of the "Humean" regularity account. However, the main arguments I put forward should go through for any acceptable "Humean" account of lawhood.

In *Counterfactuals* (Lewis 1973: 73) and "Humean Supervenience Debugged" (1994: 478), Lewis takes as a starting point a short note written by Frank Ramsey in 1928[2]. Here is how to formulate Lewis' view. Consider a true deductive system in which the general claims that represent laws of nature appear as a set of true sentences that is deductively closed and whose non-logical vocabulary contains only predicates that express *perfectly natural* properties[3]. The laws of nature will belong to all the axiom systems with a best combination of simplicity and strength[4] (and *fit*).

We can think of a true deductive system as a set of true sentences T that is deductively closed. To say that T is deductively closed is to say that every sentence that can be deduced from T is itself a member of T. The sentences that are logical consequences of T are its theorems. And there are many ways in which systems can be axiomatised. Some systems are stronger than others, in the sense that they have more information content. Some true deductive systems can be axiomatised simpler than others, in the sense that they have fewer axioms. We can have, for example, a

[2]Lewis's restatement of Ramsey's passage asserts that "a contingent generalisation 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" (Lewis 1973: 73).

[3]I have introduced the term "perfectly natural properties" in the formulation of Lewis' BSA because it is needed in order to solve a trivialisation problem (presented in the appendix of this chapter), but it is important to note that the term was not present in Lewis' original formulation.

[4]This looks like a use/mention confusion but it is almost certainly harmless. I shall try to avoid confusing the two but occasionally for ease of expression I will follow Lewis in talking in the formal mode when the claim is really one in the material mode.

very strong system that is the conjunction of all true sentences. But the complexity of such a system would not make it useful. On the other hand, we can have a less complex and simple system, only with some mathematical and logical truths. But this system would not allow us to deduce important regularities about the world. The laws of nature will belong to all the systems with a best combination of these virtues, simplicity and strength.

Notice two important features of this view. First, laws supervene on the particular matters of fact. This is so because laws merely summarise facts. So, as to potency, laws do not "govern" the world, they are just special regularities that encompass a good many other regularities. The particular matters of fact determine the laws of nature in the sense that if the laws of nature are different, that is because the facts are different. Because of this, the BSA preserves the alleged intuition that the laws of nature are metaphysically contingent, at least so long as it is metaphysically contingent what the facts are.

Given this brief description we can look at how the BSA deals with questions of whether L is compatible with contingencies.

2.1.1 Extent

Does L cover everything that happens? And what does it mean to say that L covers everything? Following Earman, one might formulate the question as follows: Do laws have an unrestricted range in space and time? (Earman 1978: 174). As Earman points out, to deny that laws have an unrestricted range in space and time boils down to saying that there is "a region of space-time R_0 such that, as far as L is concerned, 'anything goes' in R_0." (Earman 1978: 174)[5]. More precisely, where $\mathcal{M}$ denotes the set of all models of the putative law sentences, this may be formulated as the question of whether claims representing the complete and correct set of laws

[5]I use this formulation because someone may be familiar with it. But - as Cartwright says - there is no reason to assume that nature thinks in terms of space-time regions rather than, as in her view, in terms of what features obtain. For instance, her rendering of boundaries on the range of a theory T is roughly this: Those instances of effect E that T covers are the instances for which some or all of the causes of E fall under concepts available in T.

L satisfies the following condition, where $A \mid B$ stands for B is restricted to A and $\approx$ stands for model isomorphism:

(U) There is no non-empty, proper sub-region R_0 of space-time such that for any $M \in \mathcal{M}$, there is an $M' \in \mathcal{M}$ where $M' \vDash L$ and $M' \mid R_0 \approx M \mid R_0$

This condition states that L is valid on a model that is not restricted to some spatio-temporal region, that is, L is "universal". Given the BSA there is at least some motivation for thinking that the laws of nature should be "universal". If the range of the axioms (or theorems) of the best deductive systems were limited to some spatio-temporal region, then one would expect more axioms to summarise the whole history of the world. That is, one would need more axioms to cover all spatio-temporal regions. But if the range of the axioms is not limited, then one can naturally expect fewer axioms to summarise all the particular matters of fact. Furthermore, this will not reduce the system's informativeness, since the axioms now are not restricted to some spatio-temporal region.

Earman is right in pointing out that the BSA explains why we might expect the laws of nature to be universal. However, he is also right in calling our attention to the fact that there is no *a priori* guarantee that the laws of nature according to the BSA will satisfy (U) (Earman 1978: 180). In fact, consider a scenario in which our world is quite unruly and chaotic outside a given range. If this is so, adding piecemeal information about what happens there to any set of axioms may increase informativeness at too great a cost to simplicity. So the laws may be limited in extent.

2.1.2 Permissiveness

Within its domain, under the BSA, does the correct L (plus relevant initial or boundary conditions) always single out a unique outcome? In order to answer this question one needs to bear in mind the main motivation behind Humeanism about laws. The world is void of modalities – no causings, no necessitatings, no probabilifyings; the world is nothing but a mosaic of occurent events. Laws summarise what happens in this mosaic, rather than "governing" what the particular matters of fact are.

If L is deterministic, given $P_{t0}(L, P)$, L admits only one outcome. But there is nothing in the Humean motivation that makes determinism natural. The best summary may be provided by purely probabilistic laws or by laws that constrain outcomes to a given set but do not choose among them nor lay a probability over them. As Beebee points out (Beebee 2000: 575) whether the world is best axiomatised under deterministic laws depends on how regular the world is. The world can be modality free and still irregular enough to be summarised best by non-deterministic laws.

2.1.3 Reliability

So as not to muddle together issues of extent, permissiveness, and reliability, let us consider the most difficult case for contingency in the reliability sense: where the laws have universal extent and are deterministic, allowing only one output for any relevant input. It looks at first sight as if in this case on the BSA, they must be reliable. There can be no exceptions to the correct laws. I think, however, that there is still some wiggle room and will offer a way that might be thought sympathetic to the "Humean" viewpoint that might allow for exceptions, one of which is due to Lewis himself.

Once again, and for the sake of this discussion, I propose to adapt Earman's definition of determinism in terms of possible worlds to define deterministic laws because it makes for a ready connection to the Lewis wiggle. This definition is just the same as the one I gave in chapter 2, but it is worth rehearsing it. Let $\mathcal{L}$ stand for "L is the correct set of laws", then define "Humean-deterministic" thus:

Definition 2.1.1. *Laws L are Humean-deterministic if and only if for any P that L covers and any input $P_{t0}(L, P)$ to L for P and any logically possible worlds w, w' in which $\mathcal{L}$, if w and w' agree on $P_{t0}(L, P)$ they agree on whether P obtains*

Furthermore, it is worth mentioning once more that if laws are Humean-deterministic, we have the following:

D : If L is deterministic, $\Box((\mathcal{L} \wedge P_{t0}(L, P)) \supset P)$ follows[6].

The upshot then is that laws are reliable if L is deterministic. However, Sehon (2011) has interestingly argued that a definition of deterministic laws that allows us to deduce $\Box((\mathcal{L} \wedge P_{t0}(L, P)) \supset P)$ is highly problematic. And the Humean definition of laws - at least the way I presented - definitely has this consequence.

Sehon argues that, even if the correct laws are deterministic, it should be *logically* possible that there is, for example, an interventionist God (IG) that could, say, miraculously change water into wine (Sehon 2011: 31). As Sehon says, "necessarily, if an IG exists, then it is possible that the same initial state of affairs obtains, along with the same laws of nature, and yet P is false" (Sehon 2011: 31) – i.e it is possible that $P_{t0}(L, P) \wedge \mathcal{L} \wedge \sim P$. His reasoning can be spelled out as follows (using IG to stand for "There is an interventionist God"):

1	$\Box(IG \supset \Diamond(P_{t0}(L, P) \wedge \mathcal{L} \wedge \sim P))$	Premise
2	$\Diamond IG$	Premise
3	$\Diamond(P_{t0}(L, P) \wedge \mathcal{L} \wedge \sim P)$	S4, 1, 2
4	$\sim \Box((P_{t0}(L, P) \wedge \mathcal{L}) \supset P)$	Modal Logic, 3

Notice that Sehon's main point does not depend on the premise that an interventionist God is logically possible. One might try to cast Sehon's objection merely as a call for a domain restriction: L holds everywhere that there is no interventionist God (L holds if $\sim IG$). There are two problems with such a solution, however.

First, as Sehon emphasises, even if that goes through, one might argue that determinism is incompatible, say, with the logical possibility of an interventionist demon, in the sense that, necessarily, if an interventionist demon exists, then it is possible that $P_{t0}(L, P) \wedge \mathcal{L} \wedge \sim IG \wedge \sim P$. And this would also be, according to him, an implausible consequence.

Second, "Humeans" might not like the call for domain restriction because there is no way that the domain restriction could be brought into the antecedents in the laws of nature since laws are supposed to involve only occurent features, and

[6]Please see chapter 1 for a demonstration of D.

God's intervening does not seem a good candidate for an occurent feature on any "Humean" account of "occurent" I am entertaining here[7].

So, Sehon's main worry is not about the logical possibility of an interventionist God, nor about the logical possibility of a demon in particular. It is about the logical possibility of the laws of nature being violated. Thus, Sehon urges, exceptions to what L (and $P_{t0}(L, P)$) say should happen should be possible even if determinism is true, precisely because it must be logically possible to violate the laws. And if the BSA does not accommodate that, there must be something wrong with the BSA as an account of laws.

In what follows, I will show how a Lewisian might reply to this argument, showing that the BSA may be consistent with assuming that the correct laws are deterministic and yet can be violated, at least in a sense. The task then is to show that these two propositions are consistent:

P: The correct laws L are deterministic.

Q: It is possible to violate (the correct laws) L.

The first strategy is to hedge on P, using Lewis's own notion of soft determinism, which is supposed to allow a sense in which agents are able to do things such that, if they were to do them, what L says happens would not happen (Lewis 1981: 114).

Let us assume the truth of P and thus of D, so that some statement about the distant past, $P_{t0}(L, P)$, and L logically imply, for instance, P: "Agent a did not raise her hand". What if a had raised her hand? There are three options:

1. If a had raised her hand, contradictions would have been true.

2. If a had raised her hand, $P_{t0}(L, P)$ would be false.

3. If a had raised her hand, $\mathcal{L}$ would be false.

Someone like Lewis will naturally reject option 1. Even if the agent had raised her hand, contradictions would not have been true. Lewis also denies 2. Even if

[7] A domain restriction in this case seems to make law claims tautological, which they should not be for the Humean: "As are regularly associated with Bs except when they are not".

the agent had raised her hand, the past would still be the same, so $P_{t0}(L, P)$ would still be true (Lewis 1979). Thus, if we want to say that the correct set of laws L is deterministic and sometimes we are able to act otherwise, the only option remaining consistent with Lewis's viewpoint is 3. Thus, given $P_{t0}(L, P)$ and D, $\sim P$ implies $\mathcal{L}$ is false. Yet, we are supposed to be arguing that L are the correct laws. How is that possible? Following Lewis the clue is: correct in what worlds?

To see how this works we need to draw a distinction between two senses in which one can violate a law:

Weak sense: An agent is able to do something such that, if she were to do it, a law would be violated.

Wondrous sense: An agent is able to do something such that, if she were to do it, a law would be violated and this law would be of the actual world.

For example, in the weak sense, if the agent were to have raised her hand (i.e. I assume she did not raise her hand in the actual world), contrary to what L says, then L would have been violated before the hand raising. To use Lewis's phrase, a "divergence miracle" would have happened before that, that is, there would be a violation of the laws of nature that hold at our actual world, and this violation would not be caused by a's action[8]. Note that to say that there is a violation of the laws of nature in the weak sense is not to say that the violated laws are the laws of the same world where they are violated. The term "miracle" is used to express a relation between different possible worlds. As Lewis says, "a miracle at w_1, relative to w_0, is a violation at w_1 of the laws of w_0, which are at best the almost-laws of w_1" (Lewis 1979: 469). So with a divergent miracle in our actual world, whose laws are the "almost" laws of a nearest world where L is not violated, we can violate the correct laws of that nearby world. Or vice versa. Now, if by "violating a law" we mean the weak sense where what we violate is an "almost law", not a real law, of our world, then it seems agents may be able to violate laws that are deterministic.

[8]This, however, is contentious. Lewis' compatibilist view will be discussed in much more detail in chapter 4.

But what if by "violating the laws of nature" Sehon means what I call the "wondrous sense"? The wondrous sense is the one in which the laws that are violated in the actual world are the laws of the actual world. This seems to be what Sehon has in mind when he says that, if IG, then it is possible that we have the same laws, the same past, and yet P is false. Nonetheless, if by "violating a law" Sehon means the wondrous sense, then someone like Lewis will deny that it is logically possible to violate a law in the wondrous sense. This is so because, as Lewis says, "any genuine law is at least an absolutely unbroken regularity" (Lewis 1981: 114). Given the BSA, it is clear why we cannot violate laws in the wondrous sense. Suppose it is a law that no object moves faster than light. If someone were to throw an object that moves faster than light, then that law would not be true. Since Lewis's "Humean" laws are true regularities, if it is a fact that a certain stone moves faster than light, then it cannot be a true regularity that no objects travel faster than light[9].

2.1.4 Potency and Free Will

Do the things that L speaks about happen on account of L? Or, for instance, merely in accord with L? Perhaps this is the least problematic question to answer according

[9]Cartwright has suggested that "Humeans" might however, consistent with the commitment that there are only the occurent facts of which laws are summaries, take a more instrumentalist line. The best summaries may not be required to be true, especially if this brings about a big gain in simplicity. They could admit of exceptions but be right most of the time. Or they could be wrong most all the time yet still very nearly right most, even all, the time. This is like William Wimsatt's view (1992) that laws could be templates that fit widely but in many cases not exactly. Whether admitting false claims as the correct laws is a good idea on the "Humean" view depends on what the world is really like. Cartwright (1983) has argued that high level laws in physics often get fitted to the real details of real situations only by adding ad hoc corrections. That could be because we have just missed out on the factors that support those corrections and that bring the situation genuinely under the laws. But it could be that that is just what the world is like. There is no single uniform pattern but only a template which fits widely but not exactly. If the latter is the case, the BSA can be maintained while allowing contingency in the reliability sense, so long as the demand is given up that the best summary of the facts be true. (Here it is easy to make things look simpler than they are by blurring use/mention distinctions. If laws are "false" but "nearly true", then the laws will not be facts but rather only very similar to facts.)

to the BSA. Clearly, the things that L speaks about happen merely in accord with L. This means that the BSA is a non-governing conception of laws.

Beebee (2000) has argued that a consequence of such a conception of laws is that determinism is not a threat to our ability to do otherwise. Again, assume for the sake of the argument that laws are universal in extent, impermissive for agents and reliable. If the consequence argument is sound, no one is able to do otherwise.

> I think those who are moved by this argument for incompatibilism are implicitly adopting an anti-Humean conception of laws of nature - and in particular, a conception of laws according to which laws are not just generalizations about what has happened and will happen, but rather govern what will happen. It is this thought which prompts one to think that the laws of nature place a constraint on our actions that is in some way incompatible with freedom: a constraint which forces us in some metaphysical, not-purely-logical sense to act in the way we do (Beebee 2000: 579).

I believe Beebee is right in saying that those who accept the cogency of the consequence argument are implicitly assuming an anti-Humean conception of laws[10]. If - roughly speaking - Humean laws are just the best way to summarise all past, present and future facts, it is hard to see how they could be a threat to free will, even if they turn out to be deterministic and universal. In fact, suppose this is the case. Humeans who follow Lewis will agree that we are able to break the laws in the weak sense, that is, we are able to do something such that, if we were to do it, a law would be violated because a "divergence miracle" would have happened before that. As we will see in chapter 4, this will allow us to show how Humeans may reply to the premise of the consequence argument that no one has or ever had any choice about whether $\mathcal{L}$ is true.

However, in order to show that, I would like to give some plausible interpretation of "**N**", which is something I will do only in chapter 3. Furthermore, although Humeans agree on the weak sense, there is a disagreement about whether "divergence

[10]However, I am not sure that incompatibilists need to accept a *governing* view of laws. In chapter 5 I show that the premise that the laws of nature are not "up to us" is necessarily true if the dispositionalist account is correct. And one can follow Heather Demarest in combining anti-Humean properties with Humean laws into a *Potency-Best System Account of Laws* (Demarest 2017).

miracles" can be actional-events or caused by some actional-events. Lewis thinks they cannot, whereas Beebee thinks they can. It seems to me that both views are coherent, and which one is right will depend on what control is. I shall argue in chapter 4 that if the lack of contrastive explanation (that is, if there is no explanation of why an agent performed a rather than something else) indicates lack of control, then free actions cannot be divergence miracles.

2.2 Laws as relations among universals

Fred Dretske (1977), Michael Tooley (1987), and David Armstrong (1983) developed a rival approach to the BSA. In what follows this presentation will focus on Armstrong's view. Laws of nature, according to Armstrong, are necessary relations among first-order universals. That is, on Armstrong's view, a law is a second-order relation between first-order universals. Suppose that all F's are G's and that the laws of nature ensure this. F-ness and G-ness are taken to be first-order universals. Armstrong states that a second-order contingent relation holds between these two universals. He labels this relation as "nomic necessitation" and he uses "N" to refer to it - (not to be confused with the "**N**" in the consequence and *Mind* arguments though!). Armstrong symbolises the relation of necessitation between F and G as "N(F,G)". He also claims that the holding of N entails the corresponding generalisation. So, the second order-relation N between the first order universals F and G, "$N(F, G)$", entails "All F's are G's".

On the traditional Armstrong/Tooley/Dretske view it seems that laws are reliable - what they say goes, goes[11]. At least this is the case under the assumption at the core of the view that the relations that obtain between universals make true the corresponding relations between instantiations of those universals in the real world; what happens in the empirical world depends on and must be in accord with what relations hold among universals. This also ensures that laws are powerful - things

[11]It is important to note, however, that Armstrong claims that laws can be "oaken" in the sense that the entailment of $N(F, G)$ to "all Fs are Gs" does hold for all deterministic laws. Oaken laws contrast with "iron" laws where this entailment holds, and the paragraph above concerns iron laws.

happen because they say so. So potency is assured as well.

Beebee has interestingly pointed out that the disagreement between the Humean view and the Dretske/Tooley/Armstrong view with respect to potency results in different ways to understand the notion of determinism. One way to understand determinism is in terms of the past and laws determining the future.

> For the Humean, the laws and the current facts determine the future facts in a purely logical way: you can deduce future facts from current facts plus the laws. And this is just because laws are, in part, facts about the future. (Beebee 2000: 578)

For the Humean view, we do not have laws if we rule out future facts, for we need all the facts, including future ones, in order to have those statements that will best summarise them all. This is why I have defined "Humean-deterministic" in that way, quantifying over *logically* possible worlds. Notwithstanding, things are different for Armstrong's account. While for the Humean laws are in part dependent on future facts, for Armstrong it is just the other way round. As Beebee points out, "the laws 'make' the future facts be the way they will be: the laws are the ontological ground of the future facts" (Beebee 2000: 578). As a result Armstrong has a much more substantial conception of what determinism is than the Humean. The main difference is that, for Armstrong, laws of nature "make" the future events, rather than being also part of them:

> Imagine Armstrong writing down everything that's true of the universe up to this moment. One of the things that will appear in his list will be the obtaining of N between various pairs of universals. And it's in the nature of N that its obtaining entails that those universals will carry on occurring together. "N(F,G)" expresses a relation that is already with us, so the future really is determined by some current feature of the universe. For the Humean, on the other hand, a complete list of everything that's true of the universe up to now entails nothing whatever about the future, since if future facts by definition are banned from the list, then so are laws of nature. (Beebee 2000: 578-9)

Remember that, as Beebee said, Armstrong takes the laws of nature to be the ontological ground of future facts. If the literature on ground has any bite, to say that A grounds B does not seem to be equivalent to saying that A entails B. For example, we may say that the proposition expressed by "snow is white" entails the proposition expressed by "either p or not-p" such that it is logically necessary that if snow is white, then either p or not-p. But this is merely because it is logically

necessary that either p or not-p. We would not say that the fact that either p or not-p is grounded on the fact that snow is white. When we say that a certain fact grounds another there seems to be something more than entailment, that is, the facts seem to be connected through something else, namely, *ground*. This suggests that an Armstrongian conception of determinism may not be seen as tantamount to the Humean one.

How should one understand the statement that the laws of nature ground future facts? One suggestion might be the following (see Fine 2012: 39, although Fine himself is not committed to it): the fact that A grounds the fact that B if and only if the fact that B obtains *in virtue of* the fact that A and it is a metaphysical necessity that if A then B.

Now, let us briefly go back to D, that is:

D: if $\mathcal{L}$ and L is deterministic, then $\Box((\mathcal{L} \wedge P_{t0}(L, P)) \supset P)$ follows.

If laws however are Humean-deterministic, then the box of $\Box((\mathcal{L} \wedge P_{t0}(L, P)) \supset P)$ should be read as logical necessity. On Armstrong's account, on the other hand, since laws of nature *ground* future facts, it is more natural to read the box as metaphysical necessity. This - I suggest - gives a straightforward reply to Sehon's objection. For

(**M**) $\Box_M((\mathcal{L} \wedge P_{t0}(L, P)) \supset P)$

and

(**L**) $\sim \Box_L ((\mathcal{L} \wedge P_{t0}(L, P)) \supset P)$

are not explicitly contradictory. This is so because we are dealing with two different sorts of modality. Suppose that we define "epistemic necessity" as follows: ϕ is epistemically necessary for a subject s if and only if $\sim\phi$ is ruled out by what s knows. And suppose that given everything Jamie knows, it is necessary that the Brazilian economy will hit a recession in the coming months. Let P stand for the proposition that the Brazilian economy will hit a recession in the coming months. Clearly, P is not logically necessary, since it is neither an axiom of any system, nor

it is derivable from a system through its rules. Let $\Box_E$ stand for epistemic necessity, and consider the following set:

(**S**)$\{\Box_E P, \sim \Box_L P\}$

Set (S) is not explicitly inconsistent, for $\Box_E P$ and $\sim \Box_L P$ are not explicitly contradictory. Similarly, it could be argued that the set consisting of (M) and (L) is not inconsistent either. If Sehon's worry is that determinism should be consistent with the *logical* possibility of laws being violated, then (M) will do the job as long as (M) is consistent with (L).

But the problem, one might say, still remains. Even if we are going to read the box as metaphysical necessity, how are we to understand metaphysical necessity? If metaphysical necessity is just logical necessity, then even if what I suggested above is appropriate, determinism is still incompatible with an IG being logically possible. So it may be that the Armstrongian conception of determinism is incompatible with an IG being logically possible. This is fair enough. But if it is really the case, it is hard to see what is the difference between Armstrong's view and the Humean view of determinism. After all, if metaphysical necessity is just logical necessity, Armstrong's understanding of $\Box((\mathcal{L} \wedge P_{t0}(L, P)) \supset P)$ would just be something like "the fact that P obtains in virtue of $P_{t0}(L, P)$ and L and it is a logical necessity that if $P_{t0}(L, P)$ and L then P". The only difference then would be the expression "in virtue of". But as Fine points out, "we may call an in-virtue claim a statement of ontological or metaphysical ground when the conditional holds of metaphysical necessity" (Fine 2012: 38). If we insist that metaphysical and logical necessities are equivalent, I do not see how their conceptions of determinism may be different. And Beebee - I think - has given a good motivation for thinking that they are different.

In any case, I am not going any further on this issue. All I wanted to show is that it is not so implausible to read the box as metaphysical necessity if one is keen on assuming Armstrong's account of laws. If one agrees with Armstrong about the nature of laws and is also willing to separate metaphysical necessity from logical necessity, this is a good answer to Sehon's challenge.

Laws as relations among universals: Extent and Permissiveness

Let us consider the questions about extent and permissiveness now. On *extent*, the issue seems open as well. Individual advocates may argue that laws govern all that happens. But that seems to be an add-on to the two assumptions that seem central to the account that, first, laws are relations between universals, and second, any instances of universals that figure in the laws must reflect in the appropriate way the relations among those universals. These do not by themselves imply that every feature that occurs in the world instances a universal that has such relations to others and hence the two do not seem to imply that everything that happens is in the purview of laws of nature. Even if one supposes that it makes no sense to think of features that do not fall under universals, there is still the issue of whether the associated universals all participate in the kinds of relations to one another that make for laws of nature.

Permissiveness may also be more open on the laws-as-relations-among-universals view than it seems at first sight. For there may be more relations among universals than just the one – labelled "N" – that is the truth-maker for the necessitation aspect of law claims. Some universals may be taller or more beautiful than others, which may be irrelevant to what happens in the world when these universals are instantiated. Even among world-guiding relations, necessitation may not be all there is. After all, the view presumably does not want to rule out that a probabilistic theory like quantum mechanics can be correct.

One way to allow for this is to keep only N and then suppose that the universal represented by the quantum state is N-related to a universal that we represent by a probability measure. Instantiation of this last seems troublesome though; moreover probability itself, as Bas van Fraassen (1980) argues, may best be seen as a modal notion. So, in keeping with the view that modalities reflect facts about universals and their relations, another idea for how to handle probabilistic laws is to assume there is another kind of modality beyond that responsible for necessity: "probabilifies", with various ways to develop this idea further. Key though is that if the universal corresponding to A probabilifies the universal corresponding to quantity Q in accord with Prob (Q = q), then instances of A will be associated with instances

of values of Q in a pattern reflecting Prob (Q = q).

This leads readily to admitting permissiveness of the kind we see in Cartwright's account of laws (cf. Cartwright & Pemberton 2013). Once more world-guiding relations are admitted than N, there seems no good reason to suppose that an even weaker modal notion than "probabilifies" may obtain, one that constrains the values Q may take when A is instantiated to a given set but which dictates no particular pattern to them. One or another in the set must be instantiated but which on any occasion is mere hap, with not even a nice probability-looking pattern to emerge in the long run.

This may, at first sight, seem counter to the universals account of laws. After all, wasn't the point to find some location for necessity? Cartwright and I think not. The point is to find a location for modality. Universals are introduced in order to enable laws to do a number of jobs. They are supposed to support counterfactuals, to explain why things happen in the orderly way they do, to justify our inductive practice. All this may require modality, but other modalities than necessity can do the jobs required. How is it on this view that the laws of nature explain that All Fs are Gs and justify our inductive practice of predicting that the next F we encounter will be G on the basis of past observations that Fs are Gs? It is because the universal associated with F is N-related to that associated with G. But it is not the N-ness of the relation that matters; it is rather the two-fold fact that this relation holds between the universals, and whatever world-guiding relations occur must be reflected in the behaviour of their instances in the empirical world. Other kinds of patterns in the world could then be equally explained and supported by other relations between universals, for instance "F probabilifies Q=q to degree p", where the p values for Qs satisfy the probability calculus; or "F ϕ-necessitates Q", which is reflected in the fact that Fs are always followed by some value or other of Q in ϕ.

2.3 Dispositions, a la Alexander Bird

So far I have mainly focused on Lewis's and Armstrong's accounts (though I cover some new best system accounts of laws in the appendix). Although Lewis and Armstrong can both be seen as figureheads for rival camps concerning the laws of nature, Bird (2005, 2007) interestingly notes that the accounts have two theses in common: they both take (i) laws of nature to be metaphysically contingent, and they both take (ii) fundamental properties to be categorical. Dispositional essentialism (DE) has emerged as an account of laws that explicitly rejects these two assumptions. First, according to DE, laws of nature are metaphysically necessary, for reasons that we will see soon (though I shall have very little to say about what is supposed to be meant by "metaphysically necessary"). Second, DE takes at least some – maybe all – fundamental, natural properties to be essentially dispositional. I will briefly discuss in this section Bird's version of DE for a concrete illustration. Similar results with respect to contingency hold for many other versions, making appropriate adjustments.

First, Bird adopts the conditional analysis of dispositions (CA). Where D is a dispositional property, $S(D)$ is a stimulus property appropriate to it and $M(D)$ is its manifestation property, (CA) may be symbolised as follows:

(CA) $Dx \leftrightarrow (S(D)x \boxdot\!\!\rightarrow M(D)x)$

As Bird points out, (CA) does not merely provide an analysis of the concept D; instead, it characterises the nature of the property D. Thus, as Bird says, (CA) is metaphysically necessary:

($\text{CA}_{\Box}$) $\Box_M(Dx \leftrightarrow (S(D)x \boxdot\!\!\rightarrow M(D)x))$

Second, DE endorses the view that at least some fundamental properties are essentially dispositional. To say that a property P is essentially dispositional is to say that, necessarily – in the metaphysical sense – to instantiate P is to possess a disposition $D(P)$ to yield the appropriate manifestation in response to an appropriate stimulus:

(DE_p) $\Box_M\ (Px \rightarrow D(P)x)$

Here is how to explain "the truth of a generalisation on the basis of the dispositional essence of a property" (Bird 2007: 46):

1	$\Box_M(Px \rightarrow (Sx \; \Box\!\!\rightarrow Mx))$	
2	$\quad Px \rightarrow (Sx \; \Box\!\!\rightarrow Mx))$	
3	$\quad\quad Px \wedge Sx$	
4	$\quad\quad Px$	$\wedge$E, 3
5	$\quad\quad Sx$	$\wedge$E, 3
6	$\quad\quad Sx \; \Box\!\!\rightarrow Mx$	$\Rightarrow$E, 2, 4
7	$\quad\quad Mx$	$E \; \Box\!\!\rightarrow$, 5, 6
8	$\quad (Px \wedge Sx) \rightarrow Mx$	$\Rightarrow$I, 3–7
9	$\quad \forall x((Px \wedge Sx) \rightarrow Mx)$	$\forall$I, 8
10	$\Box_M(\forall x((Px \wedge Sx) \rightarrow Mx))$	$\Box$I, 1, 2–9

Hence, a universal generalisation follows from $\mathrm{CA}_\Box$ and DE_p. Furthermore, since both $\mathrm{CA}_\Box$ and DE_p are metaphysically necessary, this generalisation is metaphysically necessary as well. It looks then as if any laws underwritten by dispositional properties will be totally reliable, and on Bird's view it seems that these are all the laws there.

The problem with this, though, is that (CA) is often false, Bird notes, because of the existence of finkish dispositions and antidotes[12]. However, he argues, rather than being a disadvantage for dispositionalism, this is one of its virtues, since the falsity of (CA) allows the dispositionalist to account for ceteris paribus laws. We

[12]"An object's disposition is finkish when the object loses the disposition after the occurrence of the stimulus but before the manifestation can occur and in such a way that consequently that manifestation does not occur" (Bird 2007: 25). See also Martin (1994) and Lewis (1997). Bird also points out that one cannot eliminate all counterexamples to (CA) by excluding finks (Bird, 2007: 27). "Let object x possess disposition $D(S, M)$. At a time t it receives stimulus S and so in the normal course of things, at some later time t', x manifests M" (Bird 2007: 27). An antidote or mask to $D(S, M)$ is something that "has the effect of breaking the causal chain leading to M, so that M does not in fact occur" (Bird 2007: 27) when applied before t'. The counterexamples to (CA) will be presented in chapter 5.

can just replace the left-to-right implication of (CA) by

(V*) $\forall x$(finks and antidotes to D are absent $\rightarrow ((Dx \wedge Sx) \rightarrow Mx)$).

This is how the dispositionalist can account for *ceteris paribus* laws – supposing that in all correct *ceteris paribus* laws, the conditions that are referred to in the *ceteris paribus* clause genuinely are either finks or antidotes to the disposition referred to. Conditioning on the absence of finks and antidotes gets built right into the laws themselves. Reliability, it seems is thus restored, at least for *ceteris paribus* laws where all that is missing to render the *ceteris paribus* clause explicit is reference to finks or antidotes. Moreover, Bird also argues that there is a fundamental level of laws where no finks occur and where antidotes are very unlikely (Bird 2007: 63). In that case, as above, reliability is assured by (CA), as already noted.

What then about *permissiveness*? It seems that when they speak – which seems to be whenever a dispositional property obtains and there are no finks or antidotes to it – DE laws allow only one outcome, the manifestation associated with that disposition. So DE laws seem impermissive. On the other hand, there seems to be nothing in the basic motivations for this account that implies that the manifestation must be limited to a single choice rather than a set of choices, with or without a probability over them. So impermissivenes might be an add-on for DE laws, just as it is for laws when taken as relations among universals or on the BSA.

Extent too seems to fare just the same as in the other two accounts so far surveyed, except perhaps limitations on extent are to be expected here, at least so far as the basics I have presented go. The issue is whether everything that happens is a manifestation of (some combination of) essentialist dispositions. Two ways they may not be are immediately evident. First, if not all fundamental properties are DE properties then DE laws that supervene on DE properties and their associated dispositions will not cover them[13]. Second, DE laws derived above are, as remarked,

[13]Some proponents of DE might, however, hold a mixed view according to which some fundamental properties are essentially dispositional and others are categorical, and so a DE law could connect a disposition with a categorical property. As a result, extent may be retained since laws will not supervene only on DE properties. Thanks to Walter Ott and Lydia Patton for pointing

ceteris paribus laws, which cover only situations where no finks and antidotes obtain. What happens when these do? Or – more to the point, will finks and antidotes always be constituted by essentialist dispositional properties so that what happens when they obtain is then covered by the universal generalisation that supervenes on the dispositions associated to those? If not, then DE laws will not cover everything that happens. So DE laws may well be limited in extent.

Recall, however, that Bird maintains that there is a level of fundamental dispositional properties that are not subject to finks and are seldom - if ever - subject to antidotes. Does this imply that the correct set of laws covers all that happens? Supposing we substitute "never" for "seldom", the answer is "yes", if a kind of total reductionism holds in which everything ultimately is covered by laws deriving from fundamental dispositional properties. But this kind of reductionism does not seem to follow from the basic motivating ideas of a DE account of laws. As with many of the other assumptions I have discussed, it is just an add-on.

The real issue for extent then depends on two things. First, are all properties, including those that feature in finks and antidotes, essentially dispositional? And second, are all complexes of properties – like: "P(D) and the properties that characterise antidote A to P(D) and fink F to P(D)" – themselves essentially dispositional properties and hence properties that give rise to laws that can cover every case? Even if the answer is "yes", that does not seem to be a central part of the DE view. It seems to be just add-on. (In chapter 5 I will introduce Handfield's idea that dispositional essentialism is compatible with space-invasion. And if laws have a space-invader clause, they will be limited in extent). Naturally, one might insist that DE laws are, by their nature, universal in extent, but it does not seem to be totally clear why this should be the case.

What about reliability? Again let us look at what seems to be the hardest case – where the laws are deterministic, which is where much of the current philosophy of religion and metaphysics literature focuses. As we saw before, if the correct laws L are deterministic, then $\Box((P_{t0}(L, P) \wedge \mathcal{L}) \supset P)$.

---

this out.

This is true also for Bird's account. But the main difference between Bird's view and the BSA is how they reply to Sehon's objection. If Sehon is right, then determinism should be compatible with "IG" being logically possible. However, it should be noted that, in Sehon's argument, he reads boxes and diamonds as logical necessity and possibility. Thus, his reasoning is only relevant if the box of $\Box(P_{t0}(L, P) \wedge \mathcal{L} \supset P)$ is read as logical necessity. It will be clearer if we present his reasoning again. Let $\Box_L$ and $\Diamond_L$ respectively stand for logical necessity and possibility.

1	$\Box_L(IG \supset \Diamond_L(P_{t0}(L, P) \wedge \mathcal{L} \wedge \sim P))$	Premise
2	$\Diamond_L IG$	Premise
3	$\Diamond_L(P_{t0}(L, P) \wedge \mathcal{L} \wedge \sim P)$	S4, 1, 2
4	$\sim \Box_L ((P_{t0}(L, P) \wedge \mathcal{L}) \supset P)$	def., 3

As we can see, 4 implies the contradictory of $\Box((P_{t0}(L, P) \wedge \mathcal{L}) \supset P)$ if the box is read as logical necessity.

Now, if we take the initial or boundary conditions that feed into laws to be facts about the past, which is one typical choice for them, then let us remember what Beebee says about the issue of logical necessity for "Humean' views":

> For the Humean, the laws and the current facts determine the future facts in a purely *logical way* [my emphasis]: you can deduce future facts from current facts plus the laws. And this is just because laws are, in part, facts about the future' (Beebee 2000: 578).

According to a view motivated by Humeanism about laws, it should follow from determinism that $\Box_L((P_{t0}(L, P) \wedge \mathcal{L}) \supset P)$, as indeed it does under the definition I adopted before. That is, according to the BSA on Lewis' formulation, determinism is incompatible with "IG" being logically possible as possibility is characterised by Sehon.

On the other hand, if DE is correct, then the same strategy I used for Armstrong may work here. At a first glance, it seems that determinism could be compatible with "IG" being logically possible even as characterised by Sehon. This is so because the dispositionalist needs only one genuine notion of necessity that applies to issues about what happens in the world, which is metaphysical necessity (Bird 2007: 48). And metaphysical necessity is not logical necessity (Bird thinks that logical necessity

is not even a genuine sort of necessity). As a result, the box of $\Box((P_{t0}(L, P) \wedge \mathcal{L}) \supset P)$ stands for metaphysical necessity. Now it is clear that

(L) $\Box_L((P_{t0}(L, P) \wedge \mathcal{L}) \supset P)$

and

(M) $\Box_M((P_{t0}(L, P) \wedge \mathcal{L}) \supset P)$

are not explicitly contradictory. Someone might argue that (L) and (M) are implicitly contradictory. If logical possibility entails metaphysical possibility, then one gets the contradictory of (M); and then (L) and (M) are implicitly contradictory. Nevertheless, the dispositionalist has no motivation for accepting the premise that logical possibility entails metaphysical possibility. As Scott Shalkowski writes, essentialist claims "usually involve logical possibilities that are not metaphysical possibilities (Shalkowski 2004: 61). It is logically possible that water is not H_2O, but this is not metaphysically possible. One might object that we should expect a clear explanation of what metaphysical necessity is, since Bird's account relies on it. This might be correct, but it is something that goes beyond the purposes of this discussion. And just because we need a better understanding of metaphysical necessity it does not mean that this reply to Sehon is implausible.

How though could DE reject the "logically necessary" reading of the box since I argued that that reading follows from the definition of determinism I adopt, which is not an unconventional one? It seems the trick would be to revise the definition of determinism so that it does not involve logical necessity either but only metaphysical necessity, thus:

Definition 2.3.1. *Laws L are DE-deterministic if and only if for any P that L covers and any $P_0(L, P)$ and any metaphysically possible worlds w, w' in which $\mathcal{L}$, if w and w' agree on $P_0(L, P)$ they agree on whether P obtains*

This may indeed be a reasonable move for the DE advocate to make given the view that the only modalities that should play a role in these discussions about nature and its laws and possibilities are metaphysical ones.

The second point concerns the question of free will. Lewis' view gives motivation for rejecting one of the premises of the consequence argument, namely, the premise

that the laws of nature are not up to anyone. On the other hand, it seems that those sympathetic to DE should accept this premise because they should, it seems, accept not only rule (α) but the rule (α'):

(α') $\Box_M \phi \vdash \mathbf{N}\phi$

Given that according to DE $\Box_M \mathcal{L}$, it follows that $\mathbf{N}\mathcal{L}$. Then again, before we consider the implications of DE for the problem of free will and determinism we need to have an appropriate interpretation of "**N**". I will come back to this point in chapter 5.

2.4 Final remarks

I hope to have clarified that even if laws govern and in some sense "make things happen", there is nothing in the very nature of law in any of the senses surveyed that implies that things could not happen other than the way they do consistent with the laws staying the same, nor even that probabilities need be fixed. Laws may be universal in extent and yet totally impermissive, and one may – or may not – have good independent arguments for these add-ons; but in all senses of "laws" surveyed that is just what these are: add-ons.

One surprise of this chapter is that there are a number of different forms of "contingency" that are worth distinguishing and, contrary to what one might expect, contingency is no more readily admissible, for example, in Cartwright's capacities account than on those that take laws as strong unifying regularities (BSA), as relations among universals, or as facts about dispositions of the Alexander Bird style (or as the metaphysically necessary facts about regularities that follow from these). All these equally can, but need not, allow laws to be both permissive and limited in extent.

The upshot is that the formulation of the problem of free will and determinism relies on add-ons about laws. However, I still need to discuss the following questions: (i) does the Humean view of laws allow for contingency in the free will sense? (ii) does the dispositionalist conception of laws allow for free will? In the next chapters

I shall argue that it is plausible to think so. There are two main ways to get rid of the dilemma presented in chapter 1: committing oneself to compatibilism or libertarianism. I shall sketch a view in chapter 5 that adopts a dispositionalist account of laws. On the other hand, it seems to me that adopting Humean laws is the most natural view for compatibilists. The first question to be discussed then is the one of whether determinism/indeterminism is compatible with free will. But before we can go any further on these issues, there is still the question of whether (β) is valid. In formulating the traditional problem I pointed out that two influential arguments for the incompatibility of determinism/indeterminism and free will relies on the validity of this rule. Even if we suppose that laws are universal in extent and deterministic, incompatibilism relies on another controversial assumption, namely, that rule (β) is valid. This is what I will discuss in the next chapter.

Chapter 3

The no-choice transfer rules and the Lewis-Stalnaker semantics

In chapter 1 I indicated that the problem of free will and determinism may also be presented as an argument against libertarianism (the view that incompatibilism is true and there is free will), one that rests on the view that both determinism and indeterminism are incompatible with free will[1] (cf. also Finch & Warfield 1998). While the consequence argument is often regarded as a source of justification for the incompatibility of free will and determinism, the *Mind* argument is supposed to justify the view that indeterminism also rules out the existence of free will. Both arguments however depend on a no-choice transfer principle. The idea is that "no-choice about" transfers across material conditionals: if no one has any choice about whether P is true and whether P materially implies Q, then no one has any choice about whether Q is true. That is, both arguments as presented in chapter 1 depend on the premise that rule (β) holds.

The aim of this chapter is to give some interpretation of "no-choice about" in which some no-choice transfer rule turns out to be valid, and so in which the

[1]However, the discussion in chapter 2 opens new possibilities. For example, if deterministic and universal laws are incompatible with free will, then if there is free will, either laws are not universal or not deterministic. One possibility is to say that what free will really requires is laws being limited in extent, rather than indeterministic for agents. I shall sketch such a view in chapter 5. In any case, I will assume that "indeterminism" here means "determinism is not true".

consequence argument and the *Mind* argument turn out to be valid. This will allow us to focus on the plausibility of the premises of the arguments rather than their validity. And the discussion will reveal itself much clearer when discussing the plausibility of the premises given these interpretations, especially in chapters 5 and 6. Because the discussion has focused on (β) in the context of the consequence argument, I will be mainly concerned in discussing the importance of (β)'s validity for the consequence argument.

In this chapter I discuss the back and forth of the discussion on (β) by rehearsing the main counterexamples to it. But since there is a huge discussion on the problem of free will and determinism related to (β) it will be unreasonable to consider all the objections to it. My strategy then is to focus on what I take to be the most interesting objections to (β) in a somewhat chronological order. The chapter is structured as follows. Section 3.1 gives a brief outline of the background of the discussion concerning (β) and the consequence argument. After that, I put forward two revealing counterexamples to it: namely, the counterexamples presented by David Widerker (1987) and McKay and Johnson (1996). Sections 3.5 and 3.6 deal with an incompatibilist strategy to get rid of the counterexamples, namely, Crisp and Warfield's desiderata to avoid the counterexamples to (β). After showing the problems with that strategy, I go on to prove that some (β)-like transfer rules hold on the Lewis-Stalnaker semantics for counterfactuals.

3.1 The context of the discussion on rule (β)

The consequence argument has been recognised to break the compatibilist hegemony over the problem of free will and determinism (Levy and McKenna 2009). Since Ginet (1983) and van Inwagen (1983) advanced the modal version of the consequence argument, most compatibilists replies were directed to what was considered its weakest point: rule (β). Let us take a look at the argument again.

> If determinism is true, then our acts are the consequence of laws of nature and events in the remote past. But it's not up to us what went on before we were born, and neither is it up to us what the laws of nature are. Therefore, the consequences of these things (including our present acts) are not up to us. (van Inwagen 1983: 16)

Here is the modal formulation:

(α) $\Box\phi \vdash \mathbf{N}\phi$

(β) $\mathbf{N}\phi, \mathbf{N}(\phi \supset \psi) \vdash \mathbf{N}\psi$

1	$\Box((\mathcal{L} \wedge P_{t0}(L,P)) \supset P)$	def 1.1.1
2	$\Box(\mathcal{L} \supset (P_{t0}(L,P) \supset P))$	modal logic, 1
3	$\mathbf{N}(\mathcal{L} \supset (P_{t0}(L,P) \supset P))$	α, 2
4	$\mathbf{N}\mathcal{L}$	premise
5	$\mathbf{N}P_{t0}(L,P)$	premise
6	$\mathbf{N}(P_{t0}(L,P) \supset P)$	β, 3, 4
7	$\mathbf{N}P$	β, 5, 6

Van Inwagen himself recognised that the weakest point of the modal version is rule (β) when he wrote that "the validity of (β) is [...] the most difficult of the premises of the Third Argument to defend" (van Inwagen 1983: 96).

However, despite being the weakest point of the argument, (β) is not obviously invalid. It is not easy to think of direct counterexamples to it. As Vihvelin interestingly notes (Vihvelin 2013: 159), (β) makes the difference between the consequence argument and a typical modal fallacy such as the fatalist fallacy. For example, suppose that the proposition expressed by the sentence "I raise my hand" is true. Here is a typical modal fallacy:

1. $\Box$(I raise my hand $\supset$ I raise my hand)

2. I raise my hand

   Thus,

3. $\Box$ I raise my hand

If the conclusion is true, then I necessarily raise my hand, and so I cannot do otherwise. Fortunately, the argument above is clearly invalid, for step 3 does not follow from steps 1 and 2. It would be valid if "necessarily I raise my hand" were one of its premises, but in this case we would already be assuming a thesis called

"fatalism", the thesis according to which we are not able to do anything other than we in fact do. Rule (β) seems to be, instead, more similar to the following valid inference:

1. $\Box(\phi \supset \psi)$
2. $\Box\phi$

   Thus,

3. $\Box\psi$

Nevertheless, the necessity expressed by "**N**" is neither logical nor metaphysical necessity. So, obviously, the validity of the argument-schema above does not show the validity of (β). But because they look similar it hints why it is difficult to present a direct counterexample to it. Here is an example given by van Inwagen that seems to show that (β) is intuitively valid (van Inwagen 1983: 98):

1. **N**(The sun explodes in the year 2100)
2. **N**(The sun explodes in the year 2100 $\supset$ All life on earth ends in the year 2100)

   Hence,

3. **N**(All life on earth ends in the year 2100).

The example above is *prima facie* persuasive in showing (β)'s validity. Van Inwagen further noticed that the pre-theoretical appeal of this inference rule is independent of incompatibilism. Even if incompatibilism were false, it is hard to see how agents would have a choice about the propositions expressed by "the sun explodes in the year 2100" and "if the sun explodes in the year 2100, then all life on earth ends in the year 2100". And so it does seem to follow that no one has any choice about whether all life on earth ends in the year 2100.

One of the ways to rebut the argument is in terms of the conditional analysis of abilities, and this is one of the objections van Inwagen considers in his original paper (cf. also Gallois 1977; Narveson 1977; Foley 1979; Slote 1982; Flint 1987). For example, suppose someone holds the following conditional analysis of abilities[2]:

[2]I am following the literature in understanding the conditional as a subjunctive one.

(the simple conditional analysis of ability): agent s could have done action-type a if, and only if, if s had chosen to do a, s would have done a.

One could argue that the conditional analysis is compatible with determinism; and if this is so, then even if the premises of the consequence argument are true, its conclusion will be false, so that the argument has to be invalid. Suppose, for example, that Mary has the ability to do a PhD thesis. If the conditional analysis is correct, then if Mary had chosen to do a PhD thesis, she would have done it. However, if we suppose that laws are deterministic and that they cover the actual event that Mary does not do a PhD, this will only entail that Mary does not do a PhD. It will not entail that "if Mary had chosen to finish her PhD thesis, she would have done it" is false, and thus it will not entail that she does not have the ability to do a PhD. Thus, even if the premises of the consequence argument are true, its conclusion is false. And if the argument is invalid, rule (β) must be invalid (assuming, of course, that rule (α) is undisputable).

Much has been said about the conditional analysis of abilities (Moore 1912; Austin 1956; Berofsky 2003; Lehrer 1968; and van Inwagen 1983). Here, I will present only a very simple counterexample to that simple analysis (simple indeed, but false, as Lewis would say). The point I want to stress is that an agent performing an action is not sufficient for that agent having the ability to perform it. Anthony Kenny (1975) has already made this point: "A hopeless darts player may, once in a lifetime, hit the bull, but be unable to repeat the performance because he does not have the ability to hit the bull" (Kenny 1975: 136). Let us assume Kenny is right. Let us also suppose that this hopeless darts player in fact chooses to hit the bull and let us suppose he succeeds in so doing, even though he does not have the ability to do that. According to the Lewis-Stalnaker semantics for counterfactuals (that I shall introduce very soon!), if any propositions P and Q are true, so is the counterfactual if P were the case, Q would be the case. Now, if it is true that the darts player chooses to hit the bull and if he in fact hits the bull, it is also true that if he had chosen to hit the bull, he would have done it. If the conditional analysis is right, then it follows that the hopeless darts player has the ability to hit the bull, contradicting our initial supposition.

Now, I am not sure whether this counterexample decisively refutes the simple analysis. But it is not obvious that it fails either. Naturally, there are some more sophisticated conditional analyses that avoid the counterexamples[3]. Given the problems with the simple analysis it seems unreasonable to conclude that it is persuasive in showing (β)'s invalidity. On top of that, there is a problem with such a strategy to argue against (β). For arguing against (β) by presupposing that determinism is compatible with the ability to do otherwise is problematic. After all, incompatibilists will not accept the premise that determinism and free will are compatible. Taking this into account, incompatibilists expect that the counterexamples to (β) should not presuppose compatibilism, for the plausibility of the rule seems not to depend on incompatibilism. This actually was the challenge issued by van Inwagen: show that (β) is invalid without presupposing compatibilism. It turned out that even opponents of the consequence argument, such as McKay and Johnson (1996: 113), accepted it. Following Crisp and Warfield (2000: 175) I will refer to this condition as the First Desideratum.

FIRST DESIDERATUM: Proposed counterexamples to (β) must not presuppose compatibilism.

From now on I will consider counterexamples that presumably satisfy that desideratum.

3.2 Widerker's counterexample to (β)

In chapter 1 I followed van Inwagen in saying that "**N**ϕ" stands for "ϕ and no one has or ever had any choice about whether ϕ". But I recognise that this might not be entirely clear. Here is another way to understand it. I will use the sentential operator form presented by Ginet (1983: 391). The English expression of this operator form is this: "P and it was not in agent's s power at time t to make it not the case that

[3]Some recent attempts include Michael Fara (2008) and Vihvelin (2004), also known as the new dispositionalism. But I shall not discuss them in the thesis, since the demonstration that (β) holds will go through without making substantial assumptions about abilities.

P". Inserting universal quantification on the variables s (the agent-variable) and t (the time-variable), we get the following sentential operator: "P and, for any s and any t, it was not in s's power to make it not the case that P'. This is how I shall understand $\mathbf{N}\phi$ in this section[4].

Here is the first counterexample (Widerker 1987: 38):

> Suppose that by destroying a bit of radium r before t_9, Sam prevents the emission of a subatomic particle by r at t_9. Suppose further that this is the only way by which Sam can make sure that r will not emit radiation at t_9. Finally suppose that Sam is the only sentient being that exists or ever existed. Let 'R' and 'S' stand for
>
> R: A bit of radium r emits at t_9 a subatomic particle.
>
> S: Sam destroys r before t_9.

Now consider the following instance of rule (β):

1. $\mathbf{N}{\sim}R$

2. $\mathbf{N}({\sim}R \supset S)$

[4] This is pretty much what van Inwagen meant by his operator. Although it might not be obviously clear what he meant by "having a choice about whether P is true", I think we should interpret his construal of this operator as equivalent to Ginet's one. There is a long passage in *An Essay on Free Will* which strongly supports this interpretation:

> [W]e shall do well to state the free-will thesis as a thesis about agents and propositions. I propose to do this by devising a way to describe our powers to act – and, by acting, to modify the world – as powers over the truth-values of propositions'. Consider the propositions [. . .]
>
> (a) 27 x 15 = 405;
>
> (b) Magnets attract iron;
>
> (c) Mary Queen of Scots was put to death in 1587;
>
> (d) I have never read The Teachings of Don Juan;
>
> (e) No one has ever read all of Hume 's Enquiry aloud;
>
> (f) The cup on my desk has never been broken.
>
> [T]he truth of (a)-(c) is something it is not and never has been *within my power to change* [my emphasis], though the truth of (d)-(f) is something that is within my power, or once was within my power, to change; (a)-(c) are true and I do not have, and never have had, any choice about this [my emphasis] (van Inwagen 1983: 66)..

So, it seems that van Inwagen is in fact assuming that "P and no one has, or ever had, any choice about whether P" is equivalent to Ginet's construal of $\mathbf{N}P$. I will eventually use both of them assuming they are equivalent.

3. **N**S

According to Widerker, this is an instance of rule (β) in which the premises seem true and the conclusion false. I will examine this example more carefully. Premise one says that "~R and, for any s and any t, it was not in s's power to make it the case that R". This premise is true if and only if both conjuncts are true. The first conjunct is true because, in the example, r does not emit a particle at t_9. But why is the second conjunct true? Why is not the case that Sam has the power to make it the case that r emits at t_9 a subatomic particle? Although Widerker does not give any reason in support of the second conjunct, there is an instructive passage where he says that (β) is not truth-preserving "in situations involving the occurrence of uncaused, random events, or events associated with free human actions" (Widerker 1987: 38). So, I guess his idea in this case is that R is a proposition about the occurrence of a random event, and no one can ensure that such an event will happen. Though I take the second conjunct to be true, it is worth investigating it in a bit more detail.

First of all, it is important to note that Widerker gives a slightly different reading of "**N**P". In his view, **N**P can be read as "P and it's not within s's power at t *to bring it about* [my emphasis] that P" (Widerker 1987: 38). It is not entirely clear whether Widerker's formulation boils down to the one given by Ginet or van Inwagen. But he indicates that there are two ways in which we can understand his construal. The first and most obvious one is this:

Definition 3.2.1. *It is within s's power at t to bring it about that P if, and only if, it is within s's power to cause the event that P describes.*

Now we can see why premise one is true if we assume this definition. While it is not completely clear what "random" means when applied to single events, I assume that Widerker takes "random event" and "uncaused event" as synonyms in his paper. If the proposition expressed by "a bit of radium r does not emit a subatomic particle at t_9" describes an uncaused event, then Sam cannot cause it to happen. Thus, assuming (def. 3.2.1), it is not within Sam's power to bring it about that ~P.

Premise two states that "$(\sim R \supset S)$ and, for any s and any t, it was not in s's power to bring it about that $\sim R$ and $\sim S$". The first conjunct is true because, in the example, Sam in fact destroys r before t_9, so that S is true and consequently the material conditional $\sim R \supset S$ is also true. The second conjunct is true, to use Widerker's understanding of **N** if, and only if, Sam cannot cause the events that $\sim R$ and $\sim S$ say to occur. Suppose, however, that Sam can cause these events to happen. Suppose he does not destroy r before t_9, causing the event that $\sim S$ describes. If this is the case, he cannot ensure that r will not emit a subatomic particle at t_9, for, according to the example, the only way by which Sam can make sure that r will not emit a subatomic particle at t_9 is by destroying it before t_9. So, if Sam does not destroy r before t_9, he cannot cause the event that $\sim R$ describes, contradicting our assumption that Sam can cause the events to happen. Thus, assuming (3.2.1), it is not within Sam's power to bring it about that $\sim R$.

Although the premises are true, the conclusion is nevertheless false. It is up to Sam whether he destroys r before t_9. If the premises are true and the conclusion is false, (β) is not valid and accordingly the consequence argument is not valid.

The interpretation given by (3.2.1) may not be very compelling to some philosophers, however. For it assumes, contrary to many philosophical accounts of causation (especially the so-called "Humean" accounts), that causal *relata* are not only particular events; in the present case, the relation is conceived to be between a particular event (particle decay) and an agent (Sam). Even so, this is a point that would lead us to a different question altogether. Furthermore, if (3.2.1) is not compelling, there is another way in which we can understand "to bring it about that" without assuming agent causation. Consider this:

Definition 3.2.2. *It is within s's power at t to bring it about that P if and only if there is an act-type a such that*

(i) it is within s's power at t to perform a, and

(ii) if s were to perform a, then it would be the case that $\sim P$.

Premise one is true assuming (3.2.2). In the example, there is nothing Sam can do such that, if he did it, r would emit a subatomic particle at t_9. Even if Sam had

not destroyed r before t_9, we are not entitled to conclude that r would have emitted a subatomic particle at t_9. The conclusion is false because, had Sam not destroyed r before t_9, it would have been the case that $\sim S$.

Now consider premise two. Sam is able not to destroy r before t_9. And we are supposing that this is the only way by which r does not emit a subatomic particle at t_9. So, if Sam were not to destroy r before t_9, $\sim S$ would be true, but R would also be true, and so the material conditional $\sim R \supset S$ would be vacuously true.

The conclusion, however, is false. Sam is able not to destroy r before t_9. And if he were to do that, S would be false. Thus, rule (β) is invalid.

To sum up, the counterexample presented by Widerker seems to show that (β) is invalid[5]. Additionally, it seems not to violate the first desideratum, since the counterexample does not seem to presuppose the truth of compatibilism.

3.3 Responding to the counterexample: strengthening rule (β)

Does the counterexample really show that (β) is invalid? Philosophers such as Crisp and Warfield (2000: 175) and Timothy O'Connor (1993a: 209) think that the counterexample is successful. They think it really shows that (β), at least as formulated by Ginet and van Inwagen, is invalid. Interestingly, Widerker did not only present a counterexample to the original rule, as he offered an inference rule which is immune to his original counterexample, namely,

(β-2) $\mathbf{N}\phi, \Box(\phi \supset \psi) \vdash \mathbf{N}\psi$

However, a suggestion made by McKay and Johnson (1996) may show that the counterexample is not a serious threat to the proponents of the consequence argument after all, although it may definitely refute rule (β). In this section, I will put forward an argument to show that a very simple revision of rule (β) will enable us to sidestep the counterexample.

[5]See O'Connor (1993a) for an attempt to repair rule (β). I shall not discuss O'Connor's revised rule because I believe it fails to rule out McKay and Johnson's counterexample.

The first point that I want to make is that Widerker's example works only on the assumption that laws are unreliable. I will show that, if determinism is true (in the sense that laws are deterministic, universal in extent and reliable), then the example is not successful if the first desideratum has bite. Let us take a look at the counterexample one more time:

1. $\mathbf{N}{\sim}R$

2. $\mathbf{N}({\sim}R \supset S)$

3. ${\sim}\mathbf{N}S$

All someone needs to do is to use that inference rule presented by Widerker himself, namely, (β-2).

Suppose that determinism is true. If determinism is true, then the past and the laws of nature determine a unique future. In the formulation of determinism used earlier, if laws are reliable, deterministic and universal, then they cover S. Thus, $P_{t0}(L, S)$ and $\mathcal{L}$ strictly imply the sentence "Sam destroys r before t_9".

1. $\Box((P_{t0}(L, S) \wedge \mathcal{L}) \supset S)$

We are still not discussing the question of whether the premises of the consequence argument are true. The concern here is about whether the conclusion of the argument follows from its premises. Now, if the premises are plausible, so is the following:

2. $\mathbf{N}(P_{t0}(L, S) \wedge \mathcal{L})$

(Notice that I am not saying that 2 follows from the premises of the consequence argument, but rather that if there is no reason to reject them, there is no reason to reject 2 either).

From (1) and (2), using (β)-2, we can infer

C. $\mathbf{N}S$

If the argument above is sound, it shows that Widerker's example is successful only if it assumes that laws are not deterministic, universal and reliable, that is, only if determinism in the standard sense is not true.

Surely there is no problem in rejecting determinism to show (β)'s invalidity. After all, the question of whether (β) is valid should be independent of our considerations about determinism. However, the problem of arguing against (β) by assuming the denial of determinism is that the consequence argument has determinism as an assumption in the proof. And as we have seen, if determinism is true and the first desideratum is a plausible one, then Widerker did not present a successful counterexample to (β). For, as McKay and Johnson suggested, all we need to do in order to sidestep his counterexample is to make the following claim. Rather than saying that **N**S follows from $\mathbf{N}(\sim R \supset S)$ and $\mathbf{N}\sim R$, we can just say that it follows from these premises plus the assumption of determinism. Consider the following argument:

1. $\Box((P_{t0}(L, S) \wedge \mathcal{L}) \supset S)$
2. $\mathbf{N}\sim R$
3. $\mathbf{N}(\sim R \supset S)$

   Hence,

4. **N**S

The above argument is not susceptible to Widerker's counterexample, for if the conclusion is false, premise (1) is false. Since the consequence argument is a conditional proof that assumes the truth of determinism, we can strengthen rule (β), as McKay and Johnson suggested. Let us call this rule "delta".

(Delta) Determinism is true, $\mathbf{N}(\phi \supset \psi)$, $\mathbf{N}\phi \vdash \mathbf{N}\psi$

In this new rule, determinism was added to the left of the turnstile. Now, if Widerker's example presupposes the denial of determinism, then it certainly does not show that rule delta is invalid, simply because determinism would be false, and then we would not have a case in which the premises of delta are true and its conclusion false. Because the consequence argument assumes determinism, it is unaffected by Widerker's counterexample *if* the first desideratum is a plausible one. All we need to do is to add determinism in the justification of steps (6) and (7) in the modal formulation.

Hence, the idea is that there can be no "indeterministic" counterexamples to rule delta. Although Widerker's example seems to satisfy the first desideratum, it does not satisfy another one. Following this idea, Crisp and Warfield proposed a second desideratum that counterexamples to (β) must satisfy in order to be persuasive (Crisp & Warfield 2000: 180):

SECOND DESIDERATUM: Proposed counterexamples to (β) must not presuppose the truth of indeterminism.

As they say, "any such example that does presuppose indeterminism allows incompatibilists to sidestep the example by employing Principle Delta in the Consequence Argument" (Crisp & Warfield 2000: 180). If there is an effective counterexample to (β), according to them, then it must satisfy both desiderata. McKay and Johnson, however, believe the counterexample they gave satisfies both desiderata. I will consider McKay and Johnson's counterexample to (β) now, and then I will return to discuss the plausibility of the desiderata given by Crisp and Warfield.

3.4 McKay and Johnson's counterexample

For starters, McKay and Johnson did not give a direct counterexample to (β). Instead, they argued that agglomeration is invalid, and that rules (α) and (β) allow us to generate agglomeration. The rule of agglomeration is this:

(Agglomeration) $\mathbf{N}\phi, \mathbf{N}\psi \vdash \mathbf{N}(\phi \wedge \psi)$

Now consider a fair coin and suppose that the coin is not tossed, but that someone could have tossed it. Let P abbreviate "The coin does not land heads" and let Q abbreviate "The coin does not land tails". At a first glance the premises seem true because P and Q are true and no one has the power to make it the case that not-P, as well as the power to make it the case that not-Q. For instance, consider an ordinary agent just like me. I do not have the power to make a fair coin land heads. Nor do I have the power to make it land tails. Still, I do have the power to make a fair coin land heads or tails.

Given this, it is true that "**N**(the coin does not land heads)". There is nothing that anyone can do such that, by doing it, it would render the proposition expressed by the sentence "the coin does not land heads" false. The conclusion "**N**(the coin does not land heads and the coin does not land on tails)" seems false, though. If someone had tossed the coin, then it would have landed heads or tails, and so the proposition expressed by the sentence "the coin does land tails and the coin does not land heads" would be false. Thus, the counterexample shows that agglomeration is invalid, at least if we assume definition 3.2.2.

McKay and Johnson's reasoning against the validity of (β) is spelled out as follows:

1	$\mathbf{N}P$	premise
2	$\mathbf{N}Q$	premise
3	$\Box(P \supset (Q \supset (P \wedge Q)))$	Logical truth
4	$\mathbf{N}(P \supset (Q \supset (P \wedge Q)))$	α, 3
5	$\mathbf{N}(Q \supset (P \wedge Q))$	β, 1, 4
6	$\mathbf{N}(P \wedge Q)$	β, 2, 5

As we already saw, the premises are true. Rule (α) is valid. But the conclusion is false. Therefore, in the reasoning above, what allowed us getting a false conclusion from true premises was precisely rule (β). Therefore, (β) is invalid.

Van Inwagen (2000) himself recognised the invalidity of rule (β) when he wrote that "since the validity of [(β)] entails the validity of agglomeration, the existence of a counterexample to agglomeration entails the existence of counterexamples to [(β)]" (van Inwagen 2000: 19 note 6). Curiously, however, he does not think that McKay and Johnson's counterexample to agglomeration is a counterexample to (β). Perhaps this is a too restrictive way to understand what a counterexample is. In any case, we may present a counterexample to (β) based on McKay and Johnson's case. Let "T" abbreviate "the coin is tossed" (Vihvelin 2013).

1. $\mathbf{N}(P \supset {\sim}T)$
2. $\mathbf{N}P$

Thus,

3. **N**~T

Premise one says that "($P \supset {\sim}T$) and, for any s and any t, it was not in s's power to make it the case that $P \wedge T$". The first conjunct is true because, in the example, the coin is not tossed, so that the consequent ~T is true and consequently the material conditional is also true. The second conjunct is also true. In order to have the power to make it not the case that the material conditional $P \supset {\sim}T$ holds, someone must be able to perform an action type a (for example, tossing a coin) such that, if someone were to perform a, it would be the case that $P \wedge T$. Yet, we cannot conclude that the coin would land heads, and so we cannot conclude that P would still be true. Premise 2 is true as we saw in the example to show the invalidity of agglomeration. But the conclusion is false. If someone were to toss the coin, it would be false that ~T. Thus, rule (β) is invalid.

According to McKay and Johnson, a distinctive characteristic of their argument against (β) is that it shows not only that (β) is invalid but that rule delta is invalid as well, so that the counterexample cannot be sidestepped by using rule delta. They think that particle decay is an indeterministic process, whereas coin flip is not. However, Crisp and Warfield have argued that the coin flip example does not satisfy both desiderata.

3.5 Crisp and Warfield's defence of (β)

In order to present Crisp and Warfield's response to McKay & Johnson, we need to see again the desiderata we have been considering so far:

FIRST DESIDERATUM: Proposed counterexamples to rule (β) must not presuppose compatibilism.

SECOND DESIDERATUM: Proposed counterexamples to (β) must not presuppose the truth of indeterminism.

McKay and Johnson's example starts with the assumption that there is a fair

coin and that no one has tossed this coin, but could have tossed it. Here is how Crisp and Warfield reply to it.

Either determinism is true or not. Suppose that determinism is true. According to the example, someone has the power to make a fair coin land heads or tails, that is, someone has the power to make it not the case that the coin does not land heads and does not land tails. If this is the case, then someone has the ability to make it not the case that some proposition is false even on the assumption of determinism. This (roughly) boils down to compatibilism. Since this violates the first desideratum, Crisp and Warfield argue, the counterexample is not successful. Now suppose that determinism is not true. If determinism is not true, then indeterminism is true, which violates the second desideratum. Since this violates one of the desiderata, the counterexample is not successful. Therefore, McKay and Johnson's counterexample is not successful.

Clearly, Crisp and Warfield's argument against McKay and Johnson depends on the plausibility of the two desiderata presented above. The question one should naturally ask is this: what are the reasons for accepting these two desiderata?

First, let us see the justification for the first desideratum. Crisp and Warfield's defence of it is given in a footnote where they cite van Inwagen. This is what they say: "see van Inwagen (1983) pp. 102-3 for a discussion of this important desideratum. It is worth noting that van Inwagen's critics accept this adequacy condition on counterexamples to Principle Beta" (Crisp and Warfield 2000: 175).

This defence has two parts. One of them is certainly not decisive; after all, just because some opponents of the consequence argument accept the first desideratum, it does not follow that this desideratum is a plausible one. But if they are arguing against McKay and Johnson and they accept this desideratum, well... The other part of the defence is the same as the one presented by van Inwagen in the passage they mentioned:

> [I]t would be nice to see a counter-example to (β) that did not presuppose the compatibility of free will and determinism. After all, the examples I gave in support of (β) did not presuppose the incompatibility of free will and determinism. I should think that if there are any counter-examples to (β), then some of them, at least, could be shown to be such independently of the question whether free will and determinism are compatible. (Van Inwagen 1983: 102)

Since the discussion about (β)'s validity is a discussion in the context of the problem of free will and determinism, any counterexample to this rule that presupposes compatibilism will beg the question, or so it seems. With respect to the second desideratum they say that "any such example that does presuppose indeterminism allows incompatibilists to sidestep the example by employing Principle Delta in the Consequence Argument" (Crisp & Warfield 2000: 180).

Are the reasons above good ones to accept the desiderata? I do not think so. I will give an argument from analogy to show that this sort of strategy is unacceptable, and then present a reason as to why the first desideratum is implausible.

3.6 Against Crisp and Warfield

If Crisp and Warfield are correct, they did not only show that McKay and Johnson's counterexample is unsuccessful. They showed that all possible counterexamples to (β) in the context of the consequence argument are unsuccessful. The reason is very simple. Assuming the second desideratum, one cannot present a counterexample that presupposes the truth of indeterminism. In order to show that (β) is invalid, all we have left is to present a possible situation in which determinism is true, $\mathbf{N}P$ and $\mathbf{N}(P \supset Q)$ is true, and $\mathbf{N}Q$ is false. Here is the problem: it is impossible to do this without violating the first desideratum. If $\mathbf{N}Q$ is false, then there is someone who has a choice about whether some proposition is true (or that there is someone who can make it not the case that Q). But if determinism is true, one will be presupposing compatibilism[6]. To see why this strategy is problematic, let us take a

[6]Erik Carlson (2003) has argued that Crisp and Warfield's strategy constitutes a false dilemma. Remember what van Inwagen says in defence of rule (β): "if there are any counter-examples to beta then some of them, at least, *could* [my emphases] be shown to be such independently of the question whether free will and determinism are compatible" (van Inwagen 1983: 102). The idea here is to revise the desiderata. Taking the first desideratum into account, it is not that proposed counterexamples to (β) must not presuppose compatibilism, but rather that they *could* be proposed without presupposing compatibilism. And the same goes for the second desideratum. A counterexample to (β) must be such that it could occur even if determinism were to hold, not that the counterexample actually occurs and determinism holds (cf. Carlson 2003). However, I

look at this inference rule:

$$\textbf{(C)} : \Box((P_{t0}(L, P) \wedge \mathcal{L}) \supset P), P \vdash \Diamond{\sim}P$$

Suppose that we are using rule (C) to argue for compatibilism, and suppose we add the following desideratum:

DESIDERATUM*: Proposed counterexamples to (C) must not presuppose the falsity of compatibilism.

As we can see, if we accept the desideratum above, it is impossible to present a successful counterexample to rule (C). Suppose one gives a counterexample to (C) showing that the premises are true and the conclusion is false. The proponents of rule (C) could reply to it by saying that this counterexample is not effective because it presupposes the falsity of compatibilism. If the conclusion is false, then it follows that $\Box P$ and so that $\mathbf{N}P$, which just means that no one has or ever had free will. And if no one has or ever had free will where determinism holds, compatibilism is false. Clearly, every counterexample to rule (C) will presuppose the falsity of compatibilism; it will presuppose that determinism and the free will thesis (in this case, that someone has a choice about whether some proposition is true) cannot both be true.

My point is that Crisp and Warfield's strategy is implausible for the same reason that the strategy employed above is implausible. Rule (C) is bizarre and cannot be defended by appealing to desideratum*. If a counterexample to rule (C) presupposes the truth of incompatibilism (or the falsity of compatibilism), then if Crisp and Warfield are right in denying the counterexamples to (β) then one will also be right in denying the counterexamples to (C). Sure this cannot be right.

If the argument from analogy is correct, it shows only that Crisp and Warfield's strategy is implausible, but it does not give a diagnosis of what is wrong with their strategy. Here is what is wrong, in my view, with their strategy.

The problem seems to be with the term "presuppose", which is ambiguous[7].

---

am not completely sure whether Carlson identifies the real problem with that strategy. For that reason I shall take a different route.

[7]Sean Choi (2006) has made the same point.

How does McKay and Johnson's (and also Widerker's) counterexample *presuppose* that compatibilism is true? The counterexample is not anything like the following argument: compatibilism is true, the premises of the consequence argument are true and rule (α) is valid, so it has to be the case that rule (β) is invalid. In other words, it does not seem to be anything like the first strategy based on the simple conditional analysis of abilities. So, they cannot understand "presupposing" in the sense that the truth of compatibilism is a *sufficient* condition of (β)'s invalidity.

On the other hand, suppose that they understand "presuppose" in its "ordinary" sense, that is, that compatibilism is a necessary condition of the counterexample. More precisely, say that P presupposes Q if and only if P entails Q. (I take this distinction from Oliver Wiertz's comments on Plantinga's book *Warranted Christian Belief* (2015)). In that sense, it does make sense to say that the counterexample presupposes compatibilism, for if it is successful, then compatibilism follows. Let us reformulate the desideratum in an explicit way:

DESIDERATUM: Counterexamples to rule delta must not entail compatibilism.

But why is this desideratum plausible? Understood in that way, Crisp and Warfield are asking for the impossible: namely, a counterexample to rule delta which does not entail compatibilism. So, it is not surprising that both desiderata will not be fulfilled. That is the same reason why desideratum* is an implausible one; after all it requires something that is impossible to do. If there are no good reasons to accept this desideratum, then this strategy to defend the consequence argument fails.

To sum up, both counterexamples presented have shown that there is at least some problem with the original formulation of the argument. And Crisp and Warfield's strategy seems too implausible. But this is enough for the background of the discussion. What I shall do now is to find an alternative way to interpret "**N**", or to put forward a different formulation of the consequence argument that rests on an inference rule that avoids McKay and Johnson's counterexample.

3.7 Rule (β) and the Lewis-Stalnaker semantics for counterfactuals

The overall aim of this section is to evaluate the question of whether (β) is a valid inference rule on the Lewis-Stalnaker semantics for counterfactuals. I will consider this, first, by discussing McKay and Johnson's counterexample on Lewis' semantics for counterfactuals. I show that the counterexample is successful in showing the invalidity of rule (β) on Lewis' semantics. After that, I propose to define "no-choice about" in terms of the might-counterfactual in order to prove that a new (β)-like inference rule holds on Lewis' semantics. While the original inference rule is *invalid* on Lewis' theory, it is actually valid on Stalnaker's. I show that agglomeration and the original (β) rule are in fact valid on Stalnaker's theory and give an explanation as to why this is the case. (I also show how this allows one to reply to Alex Blum's claim about the paradoxical result of McKay and Johnson's counterexample).

3.7.1 The counterfactual sufficiency interpretation

In "Van Inwagen's Consequence Argument" Michael Huemer (2000) discusses several interpretations of "no-choice about" by arguing that a successful interpretation should satisfy the following desiderata:

a) the premises of the consequence argument are true

b) the claim that "ϕ and no one has or ever had any choice about whether ϕ" is incompatible with agents having free will

c) the argument is valid.

Since the context of this discussion concerns the consequence argument, I will follow Huemer's suggestion that we need to satisfy those desiderata.

I start with what Huemer calls "the sufficiency interpretation". Remember that to say that someone has a choice about whether P is true is to say that although P is true, someone can make it not the case that P. This can be understood as follows (Huemer 2000: 529): although P is true, someone can do something that

is sufficient for $\sim P$[8]. The following interpretation of "**N**" boils down to Huemer's counterfactual sufficiency interpretation (Huemer 2000: 529-30) and also Widerker's interpretation. Let x range over agents and α range over all past, present and future action-types (cf. also Pruss 2013). Now define "$\mathbf{N}\phi$" thus:

Definition 3.7.1. *$\boldsymbol{N}\phi$ if and only if* $\phi \wedge \sim\exists x \exists \alpha [Can(x,\alpha) \wedge (Does(x,\alpha) \boxrightarrow \sim\phi)]$

I have not defined $Can(x,\alpha)$ nor $\boxrightarrow$. Of course defining "can" is a real challenge, but as we shall see the proofs I will present work given just some fairly uncontroversial assumptions about it; and the proofs go through on Lewis' theory regardless of how we define it. With respect to $\boxrightarrow$, as the title of this chapter suggests, I shall assume the Lewis-Stalnaker semantics for counterfactuals. In other words, I shall start by saying something about the meaning of counterfactuals that amounts to the common aspects of Lewis and Stalnaker approaches:

(LS) $\phi \boxrightarrow \psi$ is true in a world w if and only if ψ is true in all the worlds in which ϕ is true that are closest to w.

I also adopt the standard terminology in saying that an ϕ-world is just a world in which ϕ is true. In this sense to say that $\phi \boxrightarrow \psi$ is true in w is to say that ψ is true in all the ϕ-worlds closest to w.

3.7.2 Lewis' theory and the *might*-counterfactual

This is the standard response to the problem. Lewis' theory allows us to see that McKay and Johnson's counterexample is indeed successful in demonstrating that (β) does not hold. In order to get Lewis' theory of counterfactuals we need to examine his definition of the "might-counterfactual", that is, "if ϕ were the case, ψ might be the case". Let $\diamondrightarrow$ stand for *if... might....* Now define $\diamondrightarrow$ thus:

(Lewis' might counterfactual): $\phi \diamondrightarrow \psi$ if and only if $\sim(\phi \boxrightarrow \sim\psi)$

[8] Van Inwagen has suggested (1983, 2004) that by "sufficiency" he meant *logical* sufficiency. But this is problematic for many reasons (cf. Huemer 2000: 529-30).

Remember that McKay and Johnson's counterexample starts with the assumption that no one tossed the coin. Let P stand for the proposition expressed by "the coin does not land heads" and Q for "the coin does not land tails". What the counterexample needs to show is that $\mathbf{N}P$, $\mathbf{N}Q$ and $\sim\mathbf{N}(P \wedge Q)$. Given definition 3.7.1, what needs to be shown is this:

1. $P \wedge \sim\exists x\exists\alpha[Can(x,\alpha) \wedge (Does(x,\alpha) \boxdot\!\!\rightarrow \sim P)]$

2. $Q \wedge \sim\exists x\exists\alpha[Can(x,\alpha) \wedge (Does(x,\alpha) \boxdot\!\!\rightarrow \sim Q)]$

3. $\sim(P \wedge Q) \vee \exists x\exists\alpha[Can(x,\alpha) \wedge (Does(x,\alpha) \boxdot\!\!\rightarrow \sim(P \wedge Q))]$

So, the first conjuncts of (1) and (2) are true, since no one tossed the coin it did not land heads and it did not land tails. Consider the second conjunct of (1) and (2). Is there a way to show that they are true? Suppose they are false. If they are false, there must be an action that an agent can perform such that, if the agent were to perform it, the coin would land tails (or heads in case of premise 2). In other words, to show that they are false, we need to show that at least one of the following counterfactuals is true:

Cf1: If someone were to toss the coin, it would land tails.

Cf2: If someone were to toss the coin, it would land heads.

Notwithstanding, neither Cf1 nor Cf2 are true on Lewis's theory. Since I think Jean-Paul Vessel explained it in a sufficiently clear way in his paper "Counterfactuals for consequentialists" (2003), the following will be a mere exposition of Vessel's explanation.

The first point Vessel raises for thinking that neither Cf1 nor Cf2 are true is that their antecedents are very unspecified. All we need to do in order to satisfy the antecedents is that someone tosses the coin. But there are many ways in which one can do that. Consider Sam's ability to do it.

> [...] Sam is capable of flipping a coin in any number of ways. Variations in these influencing factors can be used to generate a plethora of different fully specified ways in which the antecedents of our two counterfactuals might be satisfied. And this implies that there are a vast number of different possible

> worlds - at least one for each fully specified way that Sam might flip a coin - in which Sam flips the demon's coin. In one possible world, Sam flips it in direction d_1, with velocity v_1, spin action s_1, and so on. In another, Sam flips it in direction d_2, with velocity v_2, spin action s_2,... In another... (Vessel 2003: 107).

But as Vessel maintains, the fact that the antecedent is extremely unspecified is not the only factor contributing to his denial of the Cf1 and Cf2 pair.

> There are plenty of examples of pairs of counterfactuals with identical underspecified antecedents and incompatible consequents such that one of the pair is, in fact, true.
>
> Consider the following counterfactuals:
>
> **Cf5:** If I were to carry my daughter Samantha towards her crib, I would drop her along the way.
>
> **Cf6:** If I were to carry my daughter Samantha towards her crib, I wouldn't drop her along the way.
>
> Cf5 and Cf6 are underspecified in much the same ways that Cf1 and Cf2 are: there is a multitude of different fully specified ways in which I can carry Samantha to her crib. But, intuitively, Cf6 seems true. Samantha is our third baby – and I've never dropped a kid. I have the ability to guarantee that whatever fully specified ways satisfy the antecedents in the closest antecedent worlds are ways that will ensure the truth of the consequent of Cf6 (Vessel 2003: 108-9).

The ability that Vessel has to guarantee the truth of the consequent of Cf6 contrasts with one's inability to ensure the truth of the consequent of either Cf1 or Cf2. So the difference is this: Sam not only lacks the ability to perform a *particular* fully specified way of satisfying the antecedent of Cf1 and Cf2, he also lacks the ability to guarantee that a fully specified way of satisfying the antecedent of Cf1 and Cf2 will lead to his desired result. Because no one has the ability to ensure that by tossing a coin it will produce the desired result, "there don't appear to be any factors that would influence the similarity relation to grant any special priority (or 'closeness') to heads-worlds over tails-worlds" (Vessel 2003: 109):

> How a Lewis-style theory for subjunctive conditionals evaluates the Cf1-Cf2 pair now becomes clear. In order for either counterfactual to be true, either all of the closest antecedent worlds under consideration must be heads-worlds or all of them must be tails-worlds. But the fact of the matter is that some such worlds are heads-worlds; others are tails-worlds. Thus, Lewis's account entails that neither of the counterfactuals is true (Vessel 2003: 110).

If this is right, then the scenario presented by McKay and Johnson is a situation in which the premises of agglomeration are true: after all, no one has the ability to do something such that, if one were to do it, the coin would land tails (heads). The conclusion of this instance of agglomeration, however, is false. Consider the following conditional:

Cf3: If someone were to toss the coin, it would land either heads or tails.

For example, while Sam lacks the ability to guarantee that a fully specified way of satisfying the antecedent of Cf1 and Cf2 will lead to his desired result, he *does* have the ability to guarantee - as Vessel says - "that whatever fully specified ways satisfy the antecedents in the closest antecedent worlds are ways that will ensure the truth of the consequent of" Cf3 (Vessel 2003: 109).

As a result, rule (β) fails given definition 3.7.1 and Lewis' theory of counterfactuals. The natural way out thus is to find a different operator, or to make use of a different inference rule (as Widerker suggested).

3.7.3 "M" and (β) on Lewis' theory

The first strategy I consider here is to use a different operator (following McKay and Johnson, and Finch and Warfield, I shall call it "**M**"). Although agglomeration does not hold on Lewis' theory, it does hold if we define "no-choice about" in terms of the might-counterfactual. This is the simplest way to avoid McKay and Johnson's counterexample. That is, even though there is nothing I can do such that, if I were to do it, the coin would land heads (tails), there is something I can do such that, if I were to do it, the coin *might* land heads (tails).

Definition 3.7.2. *$\mathbf{M}\phi$ if and only if* $\phi \wedge \sim\exists x \exists \alpha [Can(x,\alpha) \wedge (Does(x,\alpha) \diamond\!\!\rightarrow \sim\phi)]$

3.7.4 A demonstration of $\mathbf{M}\phi$, $\mathbf{M}(\phi \supset \psi) \vdash \mathbf{M}\psi$ on Lewis' theory

Although Finch and Warfield have suggested this strategy a while ago, they did not prove that $\mathbf{M}\phi$, $\mathbf{M}(\phi \supset \psi) \vdash \mathbf{M}\psi$. Given definition 3.7.2 we can demonstrate its corresponding (β)-rule on Lewis' theory.

(β-M): $\mathbf{M}\phi$, $\mathbf{M}(\phi \supset \psi) \vdash \mathbf{M}\psi$

I will employ the following inference rules (which spring from Lewis' definition of the "might-counterfactual"):

(L) $\phi \diamond\!\!\rightarrow \psi \dashv\vdash \sim(\phi \mathbin{\Box\!\!\rightarrow} \sim\psi)$

1	$\mathbf{M}\phi$	
2	$\mathbf{M}(\phi \to \psi)$	
3	$\phi \wedge \sim\exists x\exists\alpha[Can(x,\alpha) \wedge (Does(x,\alpha) \diamond\!\!\rightarrow \sim\phi)]$	def, 1
4	$(\phi \to \psi) \wedge \sim\exists x\exists\alpha[Can(x,\alpha) \wedge (Does(x,\alpha) \diamond\!\!\rightarrow \sim(\phi \to \psi))]$	def, 1
5	ψ	$\Rightarrow$E, 1, 2
6	$\quad \sim(\psi \wedge \sim\exists x\exists\alpha[Can(x,\alpha) \wedge (Does(x,\alpha) \diamond\!\!\rightarrow \sim\psi)]$	
7	$\quad \sim\psi \vee \sim\sim\exists x\exists\alpha[Can(x,\alpha) \wedge (Does(x,\alpha) \diamond\!\!\rightarrow \sim\psi)]$	DM, 6
8	$\quad \sim\sim\exists x\exists\alpha[Can(x,\alpha) \wedge (Does(x,\alpha) \diamond\!\!\rightarrow \sim\psi)$	DS, 5, 7
9	$\quad \exists x\exists\alpha[Can(x,\alpha) \wedge (Does(x,\alpha) \diamond\!\!\rightarrow \sim\psi)$	$\neg$E, 8
10	$\quad Can(s,a) \wedge (Does(s,a) \diamond\!\!\rightarrow \sim\psi)$	$\exists$E, 9
11	$\quad \sim\exists x\exists\alpha[Can(x,\alpha) \wedge Does(x,\alpha) \diamond\!\!\rightarrow \sim(\phi \to \psi))]$	$\wedge$E, 4
12	$\quad \sim[Can(s,a) \wedge (Does(s,a) \diamond\!\!\rightarrow \sim(\phi \to \psi))]$	$\forall$E, 11
13	$\quad \sim Can(s,a) \vee \sim(Does(s,a) \diamond\!\!\rightarrow \sim(\phi \to \psi))$	DM, 12
14	$\quad Can(s,a)$	$\wedge$E, 10
15	$\quad \sim(Does(s,a) \diamond\!\!\rightarrow \sim(\phi \to \psi))$	DS, 13, 14
16	$\quad Does(s,a) \mathrel{\square\!\!\rightarrow} (\phi \to \psi)$	L, 15
17	$\quad \sim[Can(s,a) \wedge (Does(s,a) \diamond\!\!\rightarrow \sim\phi)]$	$\forall$E, 3
18	$\quad \sim(Does(s,a) \diamond\!\!\rightarrow \sim\phi)$	DS, 14, 17
19	$\quad Does(s,a) \mathrel{\square\!\!\rightarrow} \phi$	L, 18
20	$\quad\quad Does(s,a)$	
21	$\quad\quad \phi$	Logic, 19, 20
22	$\quad\quad \phi \to \psi$	Logic, 16, 20
23	$\quad\quad \psi$	$\Rightarrow$E, 21, 22
24	$\quad Does(s,a) \mathrel{\square\!\!\rightarrow} \psi$	, 20–23
25	$\quad Does(s,a) \diamond\!\!\rightarrow \sim\psi$	$\wedge$E, 10
26	$\quad \sim(Does(s,a) \mathrel{\square\!\!\rightarrow} \psi)$	L, 25
27	$\quad \bot$	
28	$\psi \wedge \sim\exists x\exists\alpha[Can(x,\alpha) \wedge (Does(x,\alpha) \diamond\!\!\rightarrow \sim\psi)]$	
29	$\mathbf{M}\psi$	

We may then proceed by putting forward a consequence argument using "**M**" rather than "**N**".

One might point out that a drawback of this operator is that it makes "**M**" agglomerative, contrary to van Inwagen's original intention. That is, perhaps what we want to capture with "**M**" is the basic locution of "having a choice about whether P is true" or "to make it not the case that P". And according to our intuitive grasp of the locution it seems that the operators should not be agglomerative. There are two responses to this criticism.

First, as Lewis once noticed (1983), since the locution was introduced as a technical term, it does not really matter what it means in natural language or even what van Inwagen originally meant. What really matters is whether we can give any meaning that would make the premises of the consequence argument defensible without circularity. I guess that this is the point of having those desiderata Huemer presented. And "**M**" seems to do the trick.

Second, even if we want to capture the intuitive grasp of the locution (assuming there is one), Blum (2000) has argued that "no-choice about" as originally intended is in fact agglomerative. Blum argues for a paradoxical conclusion, for he seems to agree that McKay and Johnson's counterexample refutes rule (β) when he writes

that "**N** ought to be, and yet ought to fail to be, agglomerative" (Blum 2000: 286)[9].

One possible response is that the intuitive locution is better interpreted with Lewis' might-counterfactual. And all that McKay and Johnson's counterexample shows is that the locution is not agglomerative if interpreted with the would-counterfactual. But we can also show that "**N**" on Stalnaker's theory is i fact agglomerative, which is in line with Blum's point.

3.7.5 Stalnaker's theory and the limit assumption

Here I will show that agglomeration holds on Stalnaker's theory given some assumption about "can" and definition 3.7.1, the counterfactual sufficiency interpretation, even though it does not hold on Lewis' theory.

In order to show that agglomeration holds we need the following inference rule (Bonevac 2003: 418).

(S) $\sim(\phi \boxdot\!\!\rightarrow \psi) \vdash \phi \boxdot\!\!\rightarrow \sim\psi$

In Stalnaker's theory (S) is easy to show. It follows from what has been said

[9]Blum's argument depends on several principles about our intuitive grasp of the locution "having a choice about whether P is true". He follows Finch and Warfield when they say that they "understand one's having a choice about a truth p as one being able to act as to ensure the falsity of p" (1998: 516). Again, I think this is just tantamount to the original understanding of the operator presented by Ginet and van Inwagen. The principles are the following (Blum 2000: 285):

Principle 1: If someone has or had a choice about whether P is true, then someone has or had a choice about whether the logical consequences of P are true as well.

Principle 2: If someone has or had a choice about whether: either p is true or [that] someone has or had a choice about whether P is true, then someone has or had a choice about whether P is true.

Principle 3: If both of two humanly unavoidable truths are each humanly unavoidable then so is their disjunction.

If the principles are true, then "no-choice about" should be agglomerative. I will not reproduce Blum's argument since it consists of a demonstration of 23 steps. And I will not discuss the plausibility of the principles either. I will just give an explanation as to why he thinks "no-choice about" should and should not be agglomerative.

about the meaning of counterfactuals *and* the limit assumption, namely, that there is never more than one closest ϕ-world. If $\sim(\phi \boxright \psi)$ holds, then ψ does not hold at at least some closest ϕ-world. Let w_1 be such a world. Thus, ψ does not hold at w_1 where ϕ holds, which is to say that $\sim\psi$ and ϕ hold at w_1. Since w_1 is the only closest ϕ-world given the limit assumption, $\sim\psi$ holds at all closest ϕ-worlds, and so $\psi \boxright \sim\psi$ follows.

(S) goes in only one direction, so that $\sim(\phi \boxright \psi)$ does not follow from $\phi \boxright \sim\psi$, since $\phi \boxright \sim\psi$ and $\phi \boxright \psi$ hold in case ϕ is impossible. But it does follow if we suppose that $\Diamond\phi$.

(S-2) $\Diamond\phi, \phi \boxright \sim\psi \vdash \sim(\phi \boxright \psi)$

Last, there is a final assumption. To show that agglomeration holds we also need the assumption that if someone can perform some action-type α, then it is possible she performs α: in other words, I shall assume that $Can(x, \alpha)$ entails $\Diamond Does(x, \alpha)$.

3.7.6 Agglomeration on Stalnaker's theory

1	$\mathbf{N}\phi$	
2	$\mathbf{N}\psi$	
3	$\sim\mathbf{N}(\phi \wedge \psi)$	
4	$\phi \wedge \sim\exists x\exists\alpha[Can(x,\alpha) \wedge (Does(x,\alpha) \boxrightarrow \sim\phi)]$	def. N, 1
5	$\psi \wedge \sim\exists x\exists\alpha[Can(x,\alpha) \wedge (Does(x,\alpha) \boxrightarrow \sim\psi)]$	def. N, 2
6	$\sim[(\phi \wedge \psi) \wedge \sim\exists x\exists\alpha[Can(x,\alpha) \wedge (Does(x,\alpha) \boxrightarrow \sim(\phi \wedge \psi)]$	def. N, 3
7	$\sim(\phi \wedge \psi) \vee \exists x\exists\alpha[Can(x,\alpha) \wedge (Does(x,\alpha) \boxrightarrow \sim(\phi \wedge \psi))]$	DM, 6
8	ϕ	$\wedge$E, 4
9	ψ	$\wedge$E, 5
10	$\phi \wedge \psi$	$\wedge$I, 8, 9
11	$\exists x\exists\alpha[Can(x,\alpha) \wedge (Does(x,\alpha) \boxrightarrow \sim(\phi \wedge \psi))]$	DS, 7, 10
12	$Can(s,a) \wedge (Does(s,a) \boxrightarrow \sim(\phi \wedge \psi))$	$\exists$E, 11
13	$\sim\exists x\exists\alpha[Can(x,\alpha) \wedge (Does(x,\alpha) \boxrightarrow \sim\phi)]$	$\wedge$E, 4
14	$\sim\exists x\exists\alpha[Can(x,\alpha) \wedge (Does(x,\alpha) \boxrightarrow \sim\psi)]$	$\wedge$E, 5
15	$\sim Can(s,a) \vee \sim(Does(s,a) \boxrightarrow \sim\phi)$	Logic, 13
16	$\sim Can(s,a) \vee \sim(Does(s,a) \boxrightarrow \sim\psi)$	Logic, 14
17	$Can(s,a)$	$\wedge$E, 12
18	$\sim(Does(s,a) \boxrightarrow \sim\phi)$	DS, 15, 17
19	$\sim(Does(s,a) \boxrightarrow \sim\psi)$	DS, 16, 17
20	$Does(s,a) \boxrightarrow \phi$	S, 18
21	$Does(s,a) \boxrightarrow \psi$	S, 19
22	$Does(s,a) \boxrightarrow \sim(\phi \wedge \psi)$	$\wedge$E, 12
23	$Does(s,a) \boxrightarrow (\phi \wedge \psi)$	Logic, 20, 21
24	$\Diamond Does(s,a)$	Can, 17
25	$\sim(Does(s,a) \boxrightarrow (\phi \wedge \psi))$	S-2, 22, 24
26	$\bot$	
27	$\mathbf{N}(\phi \wedge \psi)$	$\neg$I, 3–26

The upshot is that agglomeration holds on Stalnaker's theory given a fairly plausible assumption about "can".

McKay and Johnson's counterexample seems successful because Cf1 and Cf2 are false according to Lewis's theory. However, Stalnaker's theory combined with the theory of supervaluations takes the truth-values of these counterfactuals to be indeterminate. That is, Stalnaker's theory also produces the desired result that Cf1 and Cf2 are not true. But rather than saying they are false, Stalnaker takes their truth-value as indeterminate. He already made this point clear in this passage:

> This time someone ran off with the coin before it was tossed. Having no other coin, Tweedledee and Tweedledum argue about how it would have landed if it had been flipped. Tweedledee is convinced that it would have landed heads, Tweedledum that it would have landed tails. Again, neither has a reason – they agree that the coin was a normal one and that the toss would have been fair. This time, there is little inclination to say that one of them must be right. Unless there is a story to be told about a fact that renders one or the other of the counterfactuals true, we will say that neither is. (Stalnaker 1984: 165)

On the standard account of supervaluationism, a sentence is true if it is true on all precisifications, false if it is false on all precisifications, and neither true nor false otherwise. Cf1 and Cf2 are neither true nor false on all precisifications. Thus they are neither true nor false. If truth is truth on all precisifications then supervaluationists account for validity in the following way: an argument is globally valid if and only if if the premises are true on all precisifications the conclusion is true on all precisifications. Since on Stalnaker's theory the premises of agglomeration are not true on all precisifications, we cannot say agglomeration is invalid. So, if we are sympathetic to Stalnaker's theory, we cannot assume that the counterexample is successful. The counterexample does not show a situation in which the premises are true and the conclusion false. It shows instead a situation in which the premises are indeterminate and the conclusion is false.

This result is totally in line with the premises of McKay and Johnson's counterexample being indeterminate rather than true, so that there is no situation in which the premises of agglomeration are true and the conclusion is false. But if agglomeration holds, it seems that those sympathetic to Stalnaker's theory should

not be worried about McKay and Johnson's argument after all. In fact, the second interesting result is that the original rule (β) holds on Stalnaker's theory.

1	$\mathbf{N}\phi$	
2	$\mathbf{N}(\phi \supset \psi)$	
3	$\phi \wedge \sim\exists x\exists\alpha[Can(x,\alpha) \wedge (Does(x,\alpha) \boxrightarrow \sim\phi)]$	def.3.7.1, 1
4	$(\phi \supset \psi) \wedge \sim\exists x\exists\alpha[Can(x,\alpha) \wedge (Does(x,\alpha) \boxrightarrow \sim(\phi \supset \psi))]$	def.3.7.1, 2
5	ϕ	$\wedge$E, 1
6	$\phi \supset \psi$	$\wedge$E, 2
7	ψ	$\Rightarrow$E, 5, 6
8	$Can(s,a) \wedge (Does(s,a) \boxrightarrow \sim\psi)$	
9	$Can(s,a)$	$\wedge$E, 8
10	$Does(s,a) \boxrightarrow \sim\psi$	$\wedge$E, 8
11	$\Diamond Does(s,a)$	"Can", 9
12	$\sim\exists x\exists\alpha[Can(x,\alpha) \wedge (Does(x,\alpha) \boxrightarrow \sim\phi)]$	$\wedge$E, 3
13	$\sim\exists x\exists\alpha[Can(x,\alpha) \wedge (Does(x,\alpha) \boxrightarrow \sim(\phi \supset \psi))]$	$\wedge$E, 4
14	$\forall x\forall\alpha\sim[Can(x,\alpha) \wedge (Does(x,\alpha) \boxrightarrow \sim\phi)]$	Logic, 12
15	$\sim Can(s,a) \vee \sim(Does(s,a) \boxrightarrow \sim\phi)$	$\forall$E, 14
16	$\sim Can(s,a) \vee \sim(Does(s,a) \boxrightarrow \sim(\phi \supset \psi)$	$\forall$E, 13
17	$\sim(Does(s,a) \boxrightarrow \sim\phi)$	DS, 8, 15
18	$\sim(Does(s,a) \boxrightarrow \sim(\phi \supset \psi)$	DS, 8, 16
19	$Does(s,a) \boxrightarrow \phi$	S, 17
20	$Does(s,a) \boxrightarrow (\phi \supset \psi)$	S, 18
21	$Does(s,a) \boxrightarrow \psi$	Logic, 19, 20
22	$\sim(Does(s,a) \boxrightarrow \psi)$	S-2, 10, 11
23	$\bot$	
24	$\sim[Can(s,a) \wedge (Does(s,a) \boxrightarrow \sim\psi)]$	$\neg$I, 8–23
25	$\forall x\forall\alpha\sim[Can(s,a) \wedge (Does(s,a) \boxrightarrow \sim\psi)]$	$\forall$I, 24
26	$\psi \wedge \forall x\forall\alpha\sim[Can(s,a) \wedge (Does(s,a) \boxrightarrow \sim\psi)]$	$\wedge$I, 7, 26
27	$\mathbf{N}\psi$	def. 3.7.1, 26

Now, there are some ways to block these results. One may naturally reject the

assumption about "can". But it is hard to see how someone may be able to do something if there is no possible world at which she performs it. It is certainly the case that some abilities are never exercised. But this is different from saying that, necessarily, abilities are never exercised.

If this is right, it explains how Blum reached that paradoxical conclusion. To show that we just need to focus on the first step of his demonstration, namely, that $\mathbf{N}(\mathbf{N}P \wedge \mathbf{N}Q)$ implies $\mathbf{N}(P \wedge Q)$. This is actually the case on Stalnaker's theory. Since "$\mathbf{N}$" is factive, $\mathbf{N}(\mathbf{N}P \wedge \mathbf{N}Q)$ implies $\mathbf{N}P$ and $\mathbf{N}Q$. And because agglomeration holds, it follows that $\mathbf{N}(P \wedge Q)$. But this, of course, will not work on Lewis' theory.

My suggestion is that it is plausible that the Lewis-Stalnaker semantics have different results with respect to agglomeration and (β), and this may explain Blum's neglected argument about "no-choice about" being and not being agglomerative.

If we however want to use the original interpretation in terms of the would-counterfactual, and if we do not want to make controversial assumptions such as the limit assumption on Stalnaker's theory, one strategy is to employ a different rule, namely, (β-2) as suggested by Widerker himself. Alexander Pruss (2013) has proved that (β-2) holds given the weakening rule for counterfactuals.

(Weakening): $\phi \boxrightarrow \psi, \Box(\psi \supset \gamma) \vdash \phi \boxrightarrow \gamma$

Weakening holds on Lewis' theory as well as on Stalnaker's. I will not, however, reproduce Pruss' demonstration here. The interested reader is invited to verify that (β-2) holds given the weakening rule. One option then is to formulate the consequence in terms of rule (β-2).

1	$\mathbf{N}(\mathcal{L} \wedge P_{t0}(L, P))$	
2	$\Box((P_{t0}(L, P) \wedge \mathcal{L}) \supset P)$	
3	$\mathbf{N}P$	β-2, 1, 2

The only difference is that premise one is stronger than the original formulation, but since the point of this chapter is to put forward a valid consequence argument (and, consequently, a valid *Mind* argument) the formulation is just fine (at least for now).

3.8 Final remarks

To sum up, after this long exposition of the background of the discussion, the original formulation of the consequence argument fails if we assume Lewis' theory of counterfactuals. This is not a problem since we can either formulate it in terms of rule (β-2) or in terms of "**M**". And as I have argued, given a plausible assumption about "can" the original argument goes through if we are willing to accept Stalnaker's theory. Although I have presented the first counterexamples in complete and boring detail, they allowed us to clarify the discussion with the possible-world semantics for counterfactuals.

This is fine for the purposes of showing that the consequence argument does not rest on an unjustified inference rule. I still have not shown that the premises are true. This will be done in chapters 5 and 6, but assuming Bird's dispositionalist account of laws (to argue for the premise that the laws are not up to us) and Finch's trans-temporality thesis (to argue for the premise that the past is not up to us). As we shall see in the next chapter, however, the Humean view gives reasons for thinking that one of the premises of the consequence argument is quite problematic.

Chapter 4

Local Miracle Compatibilism

Chapter 3 had two purposes: to put forward the back and forth of the discussion on rule (β) and to show that the main question is the one of whether the premises of the consequence (and *Mind*) argument are true. It is now high time we evaluated the plausibility of the premises. Chapter 2 introduced some philosophic accounts of laws and chapter 3 some interpretations of "no-choice about". Now I combine them in order to discuss the connection between laws and free will. Here, I discuss the connection between the Humean account of laws and free will, while in chapter 5 I look at the connection between dispositional essentialism and free will.

There is one observation before we move on. Although I have mainly focused so far on the modal formulation of the consequence argument, I will discuss in this chapter van Inwagen's "first formal argument" (FFA for short) instead. There are two reasons for that. First, I attempt to provide a very careful reconstruction of Lewis' reply to the consequence argument, one that was directed to FFA. Second, I discuss in the addendum Jonathan Westphal's (2012) objection to FFA. In any case, the arguments I put forward here should go through the modal formulation as well.

One important aspect of the Humean view of laws is that it does not allow for *potency*. This is so because the things that L speaks about happen merely in accord with L rather than on account of L, such as the necessitarian view. It is not surprising that a Humean conception of laws of nature will be more in line with compatibilism. Since Humean laws do not govern, they do not place a constraint on our actions. As a result Humeans will not have trouble in saying that the correct

set of laws L is deterministic and universal, and sometimes we are able to act otherwise. Of course, if we were to do otherwise, L would be broken... Remember that in discussing the question of whether laws are reliable Humeans who follow Lewis may accept the weak sense of "breaking the laws":

> **Weak sense**: An agent is able to do something such that, if she were to do it, a law would be violated.

"'That is to say,' my opponent paraphrases, 'you claim to be able to break the very laws of nature. And with so little effort! A marvelous power indeed! Can you also bend spoons?'" (Lewis 1981: 114)". In fact, this might look like a very implausible consequence of the Humean view of laws. But Lewis claimed that the weak sense is not problematic at all. And it is not even a sense in which one *is actually able* to break the laws. For the weak sense does not commit himself to saying that L would be broken *by one of his acts* or by something caused by one of his acts.

There is disagreement among Humeans about this claim. Philosophers such as Beebee (2002, 2003) see no problem in saying that agents are able to break the laws in the sense that the laws are violated or broken by our acts. Lewis, on the other hand, claimed that agents are able to do otherwise than they in fact did even if determinism is true, but denied that agents are able to break the laws of nature in that sense. This view is known as Local Miracle Compatibilism (LMC).

Beebee has argued that LMC is untenable because it fails to provide any reason as to why it is impossible for agents to break the laws. Beebee's main point is that there is no account of laws in line with LMC. Here, however, I argue that there is one *if* a certain principle is true. One strategy for showing that two theses P and Q are consistent is to show that P is consistent with another proposition R, and that the conjunction of P and R entails Q. If so, P will be consistent with Q. What I try to do is to show that Lewis' version of the best system account of lawhood (LBSA) is consistent with LMC given the principle that a freely performed action requires a contrastive explanation, an explanation of why the agent performed a rather than not-a. I will argue that LBSA is consistent with the principle of contrastive explanation and that their conjunction entails LMC, so that LBSA is

consistent with LMC.

This all requires clarification, of course. Since LMC is discussed in the context of the first formulation of the consequence argument, I start by briefly introducing FFA and showing how LMC comes up as a reply to it. Then, I present Beebee's objection as well as the back and forth of the discussion[1]. After reformulating Beebee's objection I go on to show that LMC is consistent with Lewis' BSA.

4.1 Lewis' reply to van Inwagen's first formal argument

The first formal argument starts with the following scenario:

> **JUDGE**: Let us suppose there was once a judge who had only to raise his right hand at a certain time, T, to prevent the execution of a sentence of death upon a certain criminal, such a hand-raising being the sign, according to the conventions of the judge's country, of a granting of special clemency. Let us further suppose that the judge – call him 'J' – refrained from raising his hand at that time, and that this inaction resulted in the criminal's being put to death. We may also suppose that the judge was unbound, uninjured, and free from paralysis; that he decided not to raise his hand at T only after a period of calm, rational, and relevant deliberation; that he had not been subjected to any 'pressure' to decide one way or the other about the criminal's death. (van Inwagen 1975: 190-1)

In the argument, let t_0 be filled in by some instant of time earlier than j's birth. P_0, L and P , respectively, by a true proposition about the total state of the world at t_0, the conjunction of all the correct laws of nature and a true proposition that expresses the whole state of the world at t. Now the argument (van Inwagen 1975: 191):

1. Determinism entails that the conjunction of P_0 and L entails P.

2. If j had raised his hand at t, P would be false.

[1] I shall discuss, in particular, Oakley's response (2006) and Graham's objection to Oakley (2008). However, I shall not discuss other objections to LMC, such as the one presented by Carl Ginet (1990) and Garrett Pendergraft's response to it (2010).

3. If 2 is true, then if j could have raised his hand at t, j could have rendered P false.

4. If j could have rendered P false, and if the conjunction of P_0 and L entails P, then j could have rendered the conjunction of P_0 and L false.

5. If j could have rendered the conjunction of P_0 and L false, then j could have rendered L false.

6. j could not have rendered L false.

   Therefore,

7. If determinism is true, j could not have raised his hand at t.

Lewis (1981) argues that 5 and 6 cannot both be true, and which one is not depends on how one interprets the phrase "to render a proposition false". (As Lewis correctly points out, the phrase was introduced as a technical jargon, and it does not really matter what it means in natural language. What does matter is whether one can give any definition of it in which the premises of the consequence argument are defensible without begging the question).

In order to look at Lewis' reply to the argument we need to define the notion of a law-breaking event. And in order to do that we have to define first the notion of an event falsifying a proposition.

Definition 4.1.1. *Event e falsifies a proposition ϕ iff, necessarily, if e occurs, then $\sim\phi$.*

One might wonder why the strict conditional is needed in the *definiens*[2]. Suppose we leave the necessity operator out, and consider the material conditional "if the event that Brutus saves Caesar's life occurs, then snow is not white". Since the conditional is vacuously true, it will follow that the event that Brutus saves Caesar's life falsifies the proposition that snow is white. Naturally, e may be replaced by any

[2]Thanks to Nancy Cartwright for raising this question.

arbitrary non-actual event and ϕ by any arbitrary true proposition. Thus any non-actual event would falsify any true proposition at the actual world. This problem is avoided with the strict conditional.

Now the notion of a law-breaking event.

Definition 4.1.2. *Event e is a law-breaking event relative to world w iff, necessarily, if e occurs, then ~L, where L is the proposition stating the conjunction of all the laws of nature of w.*[3]

With this in mind Lewis aims at disambiguating the phrase "to render a proposition false" by distinguishing two senses in which someone could have rendered a proposition false.

> Let us say that I could have rendered a proposition false in the weak sense iff I was able to do something such that, if I did it, the proposition would have been falsified (though not necessarily by my act, or by any event caused by my act). And let us say that I could have rendered a proposition false in the strong sense iff I was able to do something such that, if I did it, the proposition would have been falsified either by my act itself or by some event caused by my act (Lewis 1981: 120)

There is a minor complication in the passage above because Lewis defines a two-place relation symbol in order to distinguish the two senses of rendering a proposition false. However, the weak sense is formulated in terms of "the proposition would have been falsified". And the question is: what does falsify the proposition? It certainly needs to be an event. What seems to distinguish the weak sense from the strong one is that what falsifies a proposition in the strong sense is an actional event or an event caused by an actional event. Let us then define the two senses of rendering a proposition false as follows:

Definition 4.1.3. *Agent s can render a proposition ϕ false in the weak sense iff s is able to perform some action a and if she were to perform a, there would be an event e such that e occurs and, necessarily, if e occurs then ~ϕ.*

[3]Both definitions can be extracted from the following passage: "Let us say that an event would falsify a proposition iff, necessarily, if that event occurs then that proposition is false. For instance, an event consisting of a stone's flying faster than light would falsify a law. So would an act of throwing in which my hand moves faster than light. So would a divergence miracle" (Lewis 1981: 119).

Definition 4.1.4. *Agent s can render a proposition ϕ false in the strong sense iff s is able to perform an action a and if she were to perform a, there would be an event e such that e occurs and, necessarily, if e occurs then $\sim\phi$ and either e is identical to a or e is caused by a.*

I shall put forward Lewis' argument in as much detail as possible. But some scene setting is required before doing that. Remember that Lewis argues that premises 5 and 6 cannot both be true, and which one is not depends on how we interpret "to render a proposition false". I shall start with the weak sense. If we take definition (4.1.3), premise 6 is true if and only if it is not the case that j is able to perform some action a such that, if j were to perform it, there would be an event e such that e occurs and, necessarily, if e occurs, then $\sim L$. Since premise 6 is now read in terms of a counterfactual conditional I will consider in a bit more detail Lewis' theory of counterfactuals.

4.1.1 Counterfactuals and overall comparative similarity

According to Lewis:

- **Lewis' analysis**: "if ϕ were the case, then ψ would be the case" is (non-vacuously) true at a world w iff ψ is true in all the worlds in which ϕ is true that are closest to w.

On Lewis' view closeness is similarity; so the idea is that a counterfactual is true at the actual world if and only if the consequent is true in all the worlds where the antecedent is true that are *most similar* to the actual world. In *Counterfactuals* he pointed out that his theory was relying on a familiar notion of overall comparative similarity, one that we somehow do have and that we may use to compare "big, complicated, variegated things like whole people, whole cities, or even [...] possible worlds" (Lewis 1973: 92). Lewis argued that, in this view, we may expect most similar worlds (to a certain world w) to contain law-breaking events relative to w. Here is the argument.

Suppose a proposition P is false at world w_1 and suppose the counterfactual "if P were the case, then Q would be the case" is true at it. Lewis defines the term

"deterministic laws" as follows: laws L of a world w are deterministic if and only if there is no other possible world at which L is true that is exactly like w at some time, but not exactly like w at another time (Lewis 1979: 460). Consider now the most similar P-world to w_1. If L is true at the most similar P-world to w_1, then, given that laws are deterministic, since it is not exactly like w_1 at the time to which P is true, then it is not exactly like w_1 at *all* times. But this looks quite implausible. When considering the most similar P-world to w_1 we want it to be exactly like w_1 up to the time to which P is true, which will be impossible if L is true at the closest P-world. On the other hand, if we do want to say that the most similar P-world to w_1 is one that is exactly like w_1 up to the time to which P is true, then L will not be true at this world. Lewis took the latter alternative. His idea is that the most similar worlds are those exactly like w_1 up to the time to which P is true, and that differ from it by what he called a "divergence miracle", that is, a law-breaking event relative to w_1's laws that pushes the world off the track of w_1[4].

The upshot is that we might expect most similar worlds relative to a world w to contain law-breaking events relative to w's deterministic laws. However, it is not as though every law-breaking event should be allowed. In "Counterfactual dependence and time's arrow" Lewis warned us that we cannot rely too much on the "familiar notion" of comparative similarity, for it is not "any respect of similarity you can think of must enter into the balance of overall similarity with positive weight" (Lewis 1979: 466). There he gives a detailed account of the comparative similarity relation. But this is better understood as a response in the context of the following objection.

Some counterfactuals appear to be true even when a huge difference from actuality is required. Imagine a scenario where Nixon could have pressed a button

[4]Let me give a more precise characterisation of a "divergence miracle" that I adapt from Jonathan Bennett (1984: 62). If a world w is exactly like the actual world for some period ending at t, and unlike it for some period starting at t, and if the unlikeness is a result of an event e occurring in w at t and such that, necessarily, if e occurs then L (the conjunction of the actual laws) is false, then e is a divergence miracle. The dual of this notion is a convergence miracle; that is, if a world w is unlike the actual world for some period ending at t, and unlike it for some period starting at t, and if the likeness is a result of an event e occurring in w at t and such that, necessarily, if e occurs then $\sim L$, then e is a convergence miracle.

connected to other things that would have made a nuclear war unavoidable. Consider the counterfactual "if Nixon had pressed the button, there would have been a nuclear war" (Fine 1975). This is true in the scenario described. But a world at which a nuclear holocaust occurs will be a lot unlike the actual world. On the other hand, it seems that a world at which a law-breaking event prevents the nuclear war to occur just after Nixon pressing the button is much more similar to the actual. So, it seems Lewis' theory would not be able to accommodate true counterfactuals requiring huge differences from actuality.

Lewis responds to this objection by giving a more detailed account of the similarity relation. Roughly, his account allows a trade-off between some violation of the actual laws and some difference from the actual world in particular matters of fact.

> (1) It is of the first importance to avoid big, widespread, diverse violations of law.
>
> (2) It is of the second importance to maximise the spatio-temporal region throughout which perfect match of particular fact prevails.
>
> (3) It is of the third importance to avoid even small, localized, simple violations of law.
>
> (4) It is of little or no importance to secure approximate similarity of particular fact, even in matters that concern us greatly (Lewis 1979: 47–48)

A world at which Nixon presses the button and a nuclear war does not happen is one that certainly has a great spatio-temporal region of perfect match with the actual world. But this comes at a cost of big, widespread law-breaking events. These are not divergence miracles like the one that allows Nixon to press the button, thus pushing the world off the track of the actual world. Rather, it comes at a cost of a convergence miracle, a violation of the laws that puts it back on track with the actual world. Lewis claimed that we could often expect convergence miracles to require big, widespread violations of the laws, and divergence miracles to require only small violations. Notice that his claim was not an *ad hoc* manoeuvre to reply to the Nixon

argument, but – he claimed – a consequence of the asymmetry of overdetermination instead[5]. If the asymmetry of overdetermination went in the opposite temporal direction, convergence miracles would require smaller violations of the laws. Thus, the Nixon argument fails because if we consider Lewis' account of the similarity relation. When the actual laws are deterministic, the most similar worlds to it are worlds that contain a divergence miracle, but not a convergence one.

4.1.2 Weak and strong abilities

It is now high time we evaluated the objection to the consequence argument. I will start by considering premise 6. Remember that in order to show that premise 6 is false it has to be the case that j can render L false in the weak sense; that is, j is able to perform some action a and if j were to perform a, there would be an event e such that e occurs and, necessarily, if e occurs, then $\sim L$. Given compatibilism, assuming j can raise his hand, there are all sorts of worlds where j raises his hand, and we want to consider which of them are most similar to the one where j does not raise his hand (let us say the actual world). For our purposes we just need to consider two classes of worlds where j raises his hand.

Following Lewis (1979), we consider the class typified by w_1 first. Until shortly before t, the time at which j does not raise his hand, w_1 is exactly like the actual world. "The two match perfectly in every detail of particular fact, however minute. Shortly before t, however, the spatio-temporal region of perfect match comes to an end as w_1 and [the actual world] begin to diverge" (Lewis 1979: 468). The actual laws L are violated in some "simple, localized, inconspicuous way" (Lewis 1979: 468). j raises his hand and thus prevents the execution of the criminal. No convergence miracles occur at w_1.

Second, we consider the class typified by w_2. No miracles occur at w_2 and w_2 has

[5]Let a determinant of a fact be a minimal set of conditions that, together with the laws of nature, is jointly sufficient for the fact's occurrence. Lewis claims that, from the point of view of the actual world, events typically have very few earlier determinants, but very many later determinants. More precisely, for every fact at time t_0 there is a large number of distinct determinants at all times t_1 such that $t_1 > t_0$, but there is no $t_1 < t_0$ for which the fact has a large number of determinants.

the same deterministic laws as the actual world, so that L is true at w_2. However, as j raises his hand at t in w_2, the actual world and w_2 are unalike at t. Given that laws L are deterministic, w_2 is unlike the actual world at all times.

Consider now the ranking of priorities Lewis gives. Start with (1): It is of the first importance to avoid big, widespread, diverse violations of law. Both w_1 and w_2 avoid big, widespread, divergence violations of law because they do not contain convergence miracles. Now consider (2): it is of the second importance to maximise the spatio-temporal region throughout which perfect match of particular fact prevails. Because w_2 is unlike the actual world at all times, contrary to w_1, the ranking will tell us that w_1 is more similar than w_2. The latter would be more similar to the actual world if (3) were more important than (2). But since this is not the case, given Lewis' ranking we are compelled to say that w_1 is more similar to the actual world than w_2.

Now, given that w_1 is more similar to the actual than w_2, it is true that there would be an event e at the most similar world (that is, at w_1) that occurs and such that, necessarily, if e occurs, then $\sim L$. Thus, it is false that j cannot render L false in the weak sense[6].

How about the strong sense? What if by "could have rendered P false" we mean the strong sense? In that case premise 6 is read as "j is not able able to perform an action a such that, if j were to perform a, there would be an event e such that e occurs and, if e occurs, $\sim L$ and either e is identical to a or e is caused by a". If this is the meaning we give to van Inwagen's phrase, then Lewis thinks premise 6 is true. He agrees with van Inwagen that no one is able to perform law-breaking events, or to perform actions that cause law-breaking events. Lewis is committed to

- (WEAK) Agent s is able to perform some act a such that, if she were to perform a, there would be an event e such that e occurs and necessarily if e occurs, $\sim L$,

[6]Notice that the very same argument may be put forward against the modal formulation. The judge can do something (namely, raise his hand) such that, if he were to do it, $\mathcal{L}$ would be false. And if this holds $\mathbf{N}\mathcal{L}$ is not true.

but *not*

- (STRONG) Agent s is able to perform some act a such that, if she were to perform a, there would be an event e such that e occurs and necessarily if e occurs then $\sim L$ *and* either e is identical to a or caused by a.

In that case, however, Lewis thinks that premise 5 is the problematic one. The premise is that "if j could have rendered the conjunction of P_0 and L false, then j could have rendered L false". The idea is that the argument van Inwagen gives in support of 5 does not compel one to accept it. So let us take a look at what van Inwagen says first.

> (5) This premise may be defended as an instance of the following general principle, which I take to be analytic:
>
> 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.
>
> Consider, for example, the propositions expressed by
>
> The Spanish Armada was defeated in 1588.
>
> and
>
> Peter van Inwagen never visits Alaska.
>
> The conjunction of these two propositions is quite possibly true. At any rate, let us assume it is true. Given that it is true, it seems quite clear that I can render it false if and only if I can visit Alaska. If, for some reason, it is not within my power ever to visit Alaska, then I cannot render it false (van Inwagen 1975: 192-3).

Now let us take a look at what Lewis says. I will come back to discuss it later.

> that does nothing to support Premise 5 taken in the strong sense. Given that one could render false, in the strong sense, a conjunction of historical and nonhistorical propositions (and given that, as 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? The case of the traveler is useless in answering that question, since if the traveler could render the proposition about his future travels false in the weak sense, he could also render it false in the strong sense (Lewis 1981: 120-1).

What is Lewis' point here? I think there are two ways in which we can interpret the passage above and one is more charitable than the other. According to the first

one Lewis is simply saying that the consequence argument is not cogent; he is not saying that the argument is unsound, for he is just pointing out that the argument for premise 5 fails to compel one to accept it. Why? Because premise 5 is not an instance of the general principle the consequence argument appeals to. In the example R is a proposition in which if someone can render it false in the weak sense, then she can render it false in the strong sense as well. But we are not entitled to say that L is such a proposition. The principle the proponent of the consequence argument appeals to is the following:

a) where Q is a proposition before s's birth and R is a proposition such that if s can render it false in the weak sense, she can also render it false in the strong sense, then: if s can render Q and R false in the strong sense, then s can render R false in the strong sense.

However, it is not obvious that premise 5 - understood in terms of the strong sense - is an instance of the above principle, for we are not entitled to say that L is a proposition such that if someone can render it false in the weak sense, she can also render it false in the strong sense. Consider this:

b) where Q is a proposition before s's birth and R is a non-historical proposition, then: if s can render Q and R false in the strong sense, then s can render R false in the strong sense.

If (b) is true, then premise 5 is true. Even so, we need an argument to show that (b) is true. What Lewis seems to be pointing out is that van Inwagen gives an argument for (a), not (b), so that it is an *ignoratio elenchi* and thus does not justify the acceptance of premise 5.

The above interpretation contrasts with one that commits Lewis to be saying something stronger than that: that is, that the argument is unsound because premise 5 is false[7]. Lewis accepts that someone can render L false in the weak sense and accepts as something intuitively plausible that no one can render L false in the strong

[7]This seems natural if we take the "if-then" of premise 5 to have the force of "entails" (Gallois 1977), but van Inwagen tells this is not right (van Inwagen 1977: 107).

sense[8]. The problem with this interpretation is that Lewis does not give a reason as to why 5 is false. On Lewis' account of counterfactuals there is nothing requiring that the first divergence from what actually happens can never be an actional event. And even if it did require that by stipulation, it would just seem *ad hoc* for this problem. This is why I take the first interpretation to be more charitable than the second.

So, someone who follows Lewis may accept the weak thesis, and since the argument given for premise 5 does not compel one to accept it, one does not need to accept:

(5) If j could have rendered the conjunction of P_0 and L false, then j could have rendered L false.

To sum up, Lewis' strategy consists in disambiguating van Inwagen's phrase and to commit himself only to the weak thesis. The consequence argument is not cogent either way one interprets the phrase "to render a proposition false". Either one gives the meaning in terms of definition (4.1.3) or in terms of definition (4.1.4). If we assume definition (4.1.3), then premise 6 is false. If we assume (4.1.4), then premise 5 is unjustified.

4.1.3 Local Miracle Compatibilism

If one follows Lewis in considering how counterfactuals should be evaluated on the assumption of determinism, then one is open to reject the consequence argument as not cogent. But this is certainly not enough to motivate a compatibilist view that accepts (WEAK) and denies (STRONG) (Beebee 2003: 264). After all, as Fischer (1988) asks, why *exactly* should we discriminate between the two senses in which one can render a proposition false?

[8]Lewis tells us that his denial of the strong thesis is compatible with his analysis of causation. That is, where c and e are two distinct possible events, e causally depends on c iff if c occured e would occur, and if c didn't occur e would not occur. Let m be *the* divergence miracle. It is true that if I had not raised my hand, m would not have occurred. However, it is false that, if I had raised my hand, m would have occurred. It is false because, had I raised my hand, some or other miracle would have occurred, but not the miracle m in particular. See, however, Ekstrom (1998).

> This is, I think, a legitimate and important question. Whereas it is true that there is a gap in the incompatibilist's argument, the incompatibilist has at least shown that the compatibilist (of a certain sort) is committed to a distinction which might appear to be rather "fine". There is an incompleteness in the incompatibilist's argument, and thus it is open to a person to reject it, but in rejecting it, he may be committed to a distinction between claims about our abilities which is hard to explain and justify (Fischer 1988: 249, also in Beebee 2003).

The point then is that in order to motivate a view that distinguishes the two senses in which one can render a proposition false one needs to explain and justify why there is such a distinction. Let us briefly summarise LMC as the view according to which:

(W) Agents (sometimes) have the ability to render L false in the weak sense

(S) Agents never have the ability to render L false in the strong sense

(D) Laws L are deterministic[9]

are consistent.

If the previous discussion is correct Lewis did not provide us with a reason to think that (S) is true. And even if he had done that, Beebee argues, this would not have been enough to motivate LMC. Beebee contends that LMC is a flawed view if (S) is not necessarily true. The reason is that what motivates the claim that (S) is true is a conceptual claim about the nature of laws: "the laws of nature place absolute, inviolable constraints on what we are able to do" (Beebee 2003: 268). And if this is the case, the laws of nature place constraints on what we are able to do as a matter of conceptual necessity, so it cannot be that (S) is only contingently true.

I will accept for the sake of the argument that this is right. LMC is the view that (W) and (S) are consistent with determinism and that there are no possible worlds at which (S) is false. One may find this an attractive aspect of the view because it shares the incompatibilist intuition according to which agents can never perform law-breaking events; agents are not able to travel faster than the speed of light,

[9] "Deterministic" in Lewis' sense, which boils down to the standard view that laws are both deterministic and universal in extent.

violate the principle of the conservation of angular momentum, etc. Nevertheless, Beebee points out that Lewis does not provide us with any reason as to why the strong thesis is false (Beebee 2003: 268). She puts forward a counterexample to show that (S) is possibly false. If it turns out her counterexample is correct LMC will be in a *really* bad spot.

4.2 Beebee's objection to LMC

Before presenting the objection I would like to briefly consider Beebee's criticism of Lewis' understanding of a law-breaking event first. Remember that a law-breaking event for Lewis is an event e such that, necessarily, if e occurs, then $\sim L$. Beebee, however, thinks that this definition is extensionally inadequate. "Plenty of events that are manifestly law-breaking events [...] fail to satisfy Lewis's definition" (Beebee 2003: 265). Here is why.

She considers the non-actual event of her arriving at the pub at t (call it event e). At one e-world, w_1, she waits until the very last moment and then "spontaneously disappear, reappearing in the doorway of the pub a tenth of a second later" (Beebee 2003: 262). At another e-world, w_2, she decides to go to the pub 5 minutes before and arrive there several minutes later "in an entirely non-miraculous fashion" (Beebee 2003: 266). Beebee claims that e, as it occurs in w_1, is a law-breaking event relative to the actual world, but, as it occurs in w_2, is not:

> Lewis's definition fails to make e-at-w_1 a law-breaking event, because the conditions he imposes on law-breaking events are too strict [...] What is needed is a definition of a law-breaking event that is sensitive to the circumstances that obtain, in worlds where the event occurs, when (or perhaps immediately before) the event occurs. Whether an event is a law-breaking event depends not just on the nature of the event itself, but also on the circumstances under which it occurs (Beebee 2003: 266).

In order to overcome that problem by saying that the same event can be a law-breaking one relative to one world but not all the possible worlds, Beebee provides the following amended definition of a law-breaking event.

Definition 4.2.1. *Event e is a law-breaking event at world w_2, relative to world w_1, iff e, together with the circumstances under which it occurs at w_2, is incompatible*

with L, where L is the conjunction of all of w_1's laws.

Given definition (4.2.1), she goes on to attack LMC, first, by building up the following scenario:

> Suppose that determinism is true, and suppose I am at a real estate auction. My opponent has made a bid slightly above the price I had decided would be my maximum bid, but I really want the house. I have to come to a quick decision about whether or not to raise my hand, since the auctioneer's hammer is about to fall. I decide not to raise my hand (Beebee 2003: 268-9).

About this scenario, she asks us: although she did not raise her hand, was she able to raise it? Assuming local miracle compatibilism, it seems so. After all, "this is just the kind of ordinary ability that, according to local miracle compatibilism, deterministic agents generally posses" (Beebee 2003: 269). Of course, if she had raised her hand, some law-breaking event would have occurred. But since the ability to perform that action itself does look like a strong ability, nothing compels the local miracle compatibilist to deny that she has the ability to raise her hand.

At a first glance, then, the LMC proponents might accept that Beebee was able to raise her hand. And since LMC endorses determinism being compatible with the weak thesis, in that scenario, had Beebee raised her hand, a law-breaking event would have occurred. The controversy is about whether any of Beebee's actions could be a law-breaking event, that is, whether there are strong abilities. If LMC is correct, no act of us can be or cause law-breaking events. Beebee argues that the local miracle compatibilist is not entitled to make that claim. Here is the counterexample:

> Consider the (non-actual) event m, my deciding to raise my hand. There is no reason to suppose that the closest world at which I raise my hand cannot be a world where m is the divergence miracle, and hence a law-breaking event. In that case, if I have the ability to do M then that ability is a strong ability, and there is no reason to suppose that this is an ability that I do not possess. Hence there is no reason to suppose that (S) is true. And, of course, if there is no reason to suppose that (S) is in fact true, then a fortiori there is no reason to suppose that (S) is true at all possible worlds (Beebee 2003: 269).

Beebee's decision to raise her hand, together with the circumstances under which she performs this action, is incompatible with the laws of the world in which she does not raise her hand. Thus, her action is a law-breaking event, so that strong abilities are not impossible, contrary to LMC's claim. So, if (S) is possibly false, LMC is false, as long as (S) is a necessary truth (Beebee 2003: 273).

One might ask: why is the law-breaking event her decision? Consider, for example, a world w_1 at which she decides not to raise her hand, and one femtosecond before a law-breaking event occurs – say, her arm spontaneously disappears and reappears raised – so that the world diverges from the actual allowing her to raise her hand. Beebee's reason as to why this world does not seem the most similar to the actual than the one at which m is a divergence miracle is because at w_1 the divergence miracle looks much bigger than m. Remember the event of Beebee arriving at the pub at time t. The closest world to the actual is the one at which she arrives at the pub 5 minutes later: "the loss of a few minutes' perfect match of matters of particular fact is compensated by the fact that, given that loss, we only need a small miracle rather than a big one to get e to happen" (Beebee 2003: 269).

Is this criticism correct? Well, I think Beebee has a very interesting point. I agree that Lewis' account does not suffice in showing that the first divergence from what actually happens is never an actional event. Yet, I am not entirely sure that the counterexample is persuasive[10].

In any case, all that Beebee needs to show is that there is at least one possible world where the decision is the miracle. And - as I said before - I do not think that Lewis' account of counterfactuals can avoid this unless he makes the argument

[10]For one thing, it is not clear that the example of event e (say, agent b arriving at the pub at t) motivates the need for an amendment in Lewis' definition of a law-breaking event. I am not sure whether e at w_1 and e at w_2 are the same event. Clearly, w_2 is closer to the actual world than w_1 is because, as Beebee tells us, "the miracle required at w_1 [...] is much bigger". And because "a large miracle occurs at w_1 and only a small one occurs at w_2, w_2 is closer to the actual" (2003: 262). Lewis tells us that by "big miracle" he means a "multitude of little miracles, spread out and diverse" (Lewis 1979: 471). Now the main problem is this: does e at w_1 include a multitude of little miracles or not? If it does not, then it is not a big miracle. If it does, then why would one think that e at w_1 and e at w_2 are the same event? On Lewis' account of events they would be two different classes of spatio-temporal regions, and so different events. That is, something counts as an event only if it is a class of a spatio-temporal region (Lewis 1986: 244). Given this and the axiom of extentionality, we can say that for any events x and y, x and y are different if and only if there is at least one member of x that is not a member of y or there is a member of y that is not a member of x. In that case, e at $w1$ will be a class that has a multitude of little miracles as members, whereas e at $w2$ will not. Thus, they are different events.

in terms of a somewhat *ad hoc* approach. Perhaps Lewis could avoid this if the following were true: if an actional event were to be a law-breaking event, then it would involve a "multitude of little miracles, spread out and diverse" and thus would be a big miracle. Nevertheless, there is no reason to suppose that this would be the case. It could be that a law-breaking actional event were to involve just a simple violation of laws. This is why I agree that Lewis does not provide us with any reason to think that (S) is true, let alone necessarily true.

4.3 Interlude: Oakley's objection to Beebee

Shane Oakley (2006) has defended LMC from Beebee's objection by rejecting her definition of a law-breaking event. He claimed that according to Beebee's definition of a law-breaking event "any mundane counterfactual action that one could have performed in a deterministic world could be considered a law-breaking event" (Oakley 2006: 343), which is something unacceptable according to him. And if (4.2.1) is replaced by Lewis' definition, the counterexample will be avoided, for he thinks that Beebee's decision to raise her hand is not a law-breaking event according to Lewis' definition.

> Moreover, Beebee has not offered a convincing argument, other than to note that [Lewis' definition of a law-breaking event] precludes many mundane actions from being law-breaking events, which is exactly the motivation for Lewis advocating such a definition (Oakley 2006: 344)

Oakley accepts alongside Beebee that if (4.2.1) is a good definition of a law-breaking event, given determinism, it is possible for some non-actual action to be a law-breaking event. Suppose the truth of determinism and consider some arbitrary non-actual action a, e.g., raising one's hand. It could turn out that a occurs together with the circumstances C in a way that is incompatible with L; for example, the circumstances may require that in order to act several particles will have to travel faster than the speed of light. So depending on how the circumstances are, any non-actual action could be a law-breaking event relative to the actual world. Thus, it is precisely for this reason that he denies we should adopt definition (4.2.1) in order to formulate LMC.

In a nutshell, Oakley claims that Beebee's objection is an *ignoratio elenchi*, for it does not refute LMC as formulated by Lewis (since Lewis' LMC has a different understanding of a law-breaking event).

Another reason Oakley gives is that Beebee's supposed amendment definition (that is sensitive to the circumstances) is questionable. "Law-breaking event" is introduced as a technical term in the context of how counterfactuals should be evaluated. However, when evaluating counterfactuals on the assumption of determinism, the similarity relation is already taken to be fixed by the *context* (that is, the worlds are similar to one another if they agree on a large number of what the relevant interlocutors take to be their most important features). Therefore, if the term "circumstances" fixes the context of the similarity relation, then there is no need for it in the definition of a law-breaking event. So it is hard to see how there is motivation for accepting the supposed amendment in (4.2.1) as "counterfactuals events are not considered *in vacuo*." (2006: 343)[11].

On top of that it seems to me that Oakley could just point out that all that matters in the context of this discussion is whether one can give a definition of a law-breaking event in line with Lewis' ranking of overall similarity and LMC. Lewis' definition is just a stipulative definition of a law-breaking event and not a definition of the intuitive notion. One might even wonder whether there is an intuitive notion of a law-breaking event or miracle. We do seem to have the intuitive notion of a miracle, such as God changing water into wine, but this is not what is at stake in this discussion; "miracles" here are never actual events that violate actual laws, as I discussed in chapter 2 the question of whether laws of nature are *reliable*. So one may see no motivation to abandon Lewis' stipulative definition in favour of (4.2.1); after all, definition (4.2.1) does seem compatible with some actions being law-breaking events. Therefore, if Oakley is correct, LMC should be formulated in terms of Lewis' original definition of a law-breaking event, so that Beebee's counterexample is no threat.

[11]It is true that there are no counterfactuals in either Lewis' or Beebee's definitions. But the point here is that "law-breaking event" is introduced in this context to evaluate counterfactuals given determinism.

Why does Oakley believe that m is not a law-breaking event according to Lewis' definition? Unfortunately he does not give us a reason, but I think he has something like the following in mind (van Inwagen 1983: 3). Suppose that determinism is true and let P be a true proposition that the event that there is no eclipse tonight occurs. The standard way of cashing out determinism is that it has a consequence that P_0 and L entail P. Now consider the non-actual event that there is an eclipse tonight; this event is not contrary to the laws - as van Inwagen says - "for the laws of nature do not by themselves dictate when particular events such as eclipses shall occur" (van Inwagen 1983: 3), in the sense that it is not the case that if the event that there is an eclipse tonight occurs, then L is false. The event is nomologically impossible, but not contrary to the laws.

Peter Graham (2008), however, has argued that this solution does not protect the local miracle compatibilist from Beebee's objection. Although Graham agrees that in the estate auction scenario m - the event that she decides to raise her hand - is not a law-breaking event according to Lewis' definition, he says that

> there is nothing stopping Beebee from stipulating that in the real estate auction scenario, her neurophysiology is such that her choosing to raise her hand would require and involve a few particles in her brain travelling a short distance faster than the speed of light. Were she to stipulate this, then it is plausible that in the real estate auction scenario, the smallest miracle, or divergence from reality, required in order to bring about her raising her hand would be an event that is both identical to, or partly constitutive of, her choosing to raise her hand and a law-breaking event even in the sense given by [Lewis' definition]. (Graham 2008: 69)

Graham claims that this stipulation, together with Lewis' account of counterfactual, is sufficient for the truth of the counterfactual "if Beebee had raised her hand, there would have been an event e such that necessarily, if e occurs then $\sim L$ and e is identical to m". This may be right, but as far as I can see, it does not refute LMC understood as a reply to the consequence argument.

I agree that one may stipulate that the decision involves a few particles travelling faster than the speed of light. It could also be that the decision is the smallest miracle relative to the actual laws. Even so, this should not convince the local miracle compatibilist to accept that *it is up to* someone to do that. First, LMC can be understood as a thesis about what actions are up to agents to perform. It states

that it is up to agents to do things in the weak sense, that is, that would require the actual laws to be violated by a divergence miracle. On the other hand, given that thesis, we are not entitled to say that it follows that it is up to agents to perform events that violate the actual laws. What a counterexample to LMC must show is that at least someone is able to do something in the strong sense. Now let me go back to the counterexample. We can easily stipulate that the decision is the event that Beebee decides to raise her hand in a way that, necessarily, if it occurs, the actual laws are false. Call it event *b*. However, can we stipulate that someone is able to do *b*? If we stipulate that, we will just be stipulating that someone is able to perform an action in the strong sense. And this clearly begs the question[12].

Since I am saying that Graham's use of the counterexample begs the question I should at least motivate my claim by pointing out an account of begging the question that supports my point. Walter Sinnott-Armstrong's account (1999), although not uncontroversial, will do the trick. Sinnott-Armstrong's account of begging the question adopts an epistemic approach according to which begging the question "depends on whether one has the right kind of reason to believe the premise" (Sinnott-Armstrong 1999: 179). In addition, he embraces a subjective epistemic approach as opposed to an objective one. Roughly, on the subjective approach, he tells us that whether a use of an argument begs the question depends on the beliefs of the arguer or the audience. This is based on his distinction between audience justification (when the arguer tries to show the audience that the audience has a reason to believe in a proposition) and arguer justification (when the arguer tries to show the audience that the arguer herself has a reason to believe in a proposition). Whether these purposes are achieved depend on the beliefs of the audience (in case of audience justification) and the arguer's beliefs (in case of arguer justification).

So, on Sinnott-Armstrong's account, begging the question depends on the beliefs that give reason for the audience (or the arguer) to believe the premises. Further-

[12]One might object that if no one is able to do *b* then *b* cannot be an action. Well, if this is right, then the mere stipulation that *b* is an action will presuppose that someone is able to do *b* (and thus that someone is able to act in the strong sense). Thus, the mere stipulation that *b* is an action will make this use of the argument question-begging.

more, he tells us that, in order to avoid begging the question, the reasons to believe the premises must be independent of the conclusion (Sinnott-Armstrong 1999: 182). He gives then two sufficient conditions of begging the question:

> Thus, there are two sufficient conditions of begging the question: dependence on one's belief in the conclusion and dependence on one's reason to believe the conclusion. Contrapositively, to avoid begging the question one's reason to believe the premise must be independent of both (a) one's belief in the conclusion and also (b) one's reason to believe the conclusion. (Sinnott-Armstrong 1999: 183)

Although he does not say anything about what he means by the notion of dependence, he does give some instructive examples suggesting that dependence can be explained in terms of counterfactuals: the reason to believe in the premise depends on the belief in the conclusion only if if one were not to believe in the conclusion, then one would not have the reason to believe in the premise.

Suppose Graham is seeking audience justification, so that he is using the argument to show the audience (in our case, the local miracle compatibilist who accepts Lewis' definition of a law-breaking event) that someone is able to decide even though the decision is by stipulation a law-breaking event. However, the reason to believe in the premise depends on the belief in the conclusion: that is, (a) if one were not to believe in the conclusion, then one would not have the reason to believe in the premise. In other words, if one were not to believe that some law-breaking event is up to someone, then one would not have a reason to believe that Beebee has the ability to decide to raise her hand where this event is by stipulation a law-breaking event in Lewis' sense. And because no other reason is given as to why one is able to decide to act if the decision involves an event whose occurrence is sufficient to falsify L, the counterexample will have bite only if it already presupposes that someone is able to do some action in the strong sense. In that case it clearly begs the question.

In the original scenario m is just an ordinary action. It happens to be a law-breaking event because of the circumstances under which it occurs. Even if one were not to believe that someone is able to perform actions in the strong sense, one would still have a reason to believe that someone has the ability to m if that ability were to be a weak one, for the local miracle compatibilist believes that we have the ability to perform weak abilities. In Graham's case, however, it is hard to see why

someone has the ability to m if m by stipulation is not a mundane action, but rather a law-breaking event that includes a few particles travelling faster than the speed of light.

One however might object as follows. The local miracle compatibilist presupposes without argument that strong abilities are impossible. If that is the case then there is no reason to accept LMC, and there is no reason to deny that someone is able to perform a law-breaking event.

This criticism is correct, as Fischer pointed out that the local miracle compatibilist needs to motivate the distinction and explain why strong abilities are impossible. Yet, this is not necessary if the local miracle compatibilist just aims at rebutting the consequence argument. That is, she may not be trying to show the audience (the incompatibilist) that we never have strong abilities. Incompatibilists already agree with that. The local miracle compatibilist just points out that we are not entitled to think that the claim we have strong abilities follows from the claim we sometimes have weak abilities. So, in order to rebut the consequence argument as not cogent all that the local miracle compatibilist needs is to accept the weak thesis and deny that we have a good reason to think the strong one follows from it. And precisely because of that the local miracle compatibilist has no reason to accept that someone is in fact able to break the laws[13].

If Graham's reply does not go through, does it mean that Oakley's defence is a plausible one? If the purpose of LMC is just to reply to the consequence argument, yes. But it needs much more if it is to be taken as a serious compatibilist view about free will and determinism. And Oakley's strategy neither motivates the distinction between strong and weak abilities nor provides us with any reason why actional events can never be law-breaking events. Thus, his solution does nothing in responding to the problem that Beebee and Fischer pointed out.

On top of that, Beebee's counterexample is supposed to be a symptom of a bigger problem with LMC. Beebee's diagnosis of LMC's problem is that there is no account of laws in line with it. The notion of a law of nature is central in formulating

[13]In all fairness, Graham has formulated this weak version of LMC in a similar way as well.

LMC. If it happens that (W) and (S) cannot sound simultaneously plausible with any account of laws, then there is definitely a problem with it. Hence, it is crucial for LMC to get rid of this problem. This is what I shall do in the next section. My response will give motivation to think that LMC should accept the proposition that agents never have strong abilities.

4.4 Is LMC compatible with Lewis' BSA?

Let me briefly rehearse what is at stake here. LMC is the view according to which

- **(W)** Agents (sometimes) have weak abilities, that is, the ability to render a proposition false in the weak sense

  and

- **(S)** Agents never have strong abilities, that is, the ability to render a proposition false in the strong sense

are consistent with the truth of determinism and there are no worlds at which determinism is true, (W) is true and (S) is false. However, Beebee argues that strong abilities are possible, contrary to LMC's claim. So, if (S) is possibly false, LMC is false, as long as (S) is a necessary truth (Beebee 2003: 273). As far as I see Beebee's counterexample is an attempt to show a problem that comes from (W) and (S): there seems to be no account of laws of nature in which one plausibly holds both (W) and (S). And I agree that this *seems* to be the case, even though I am not entirely sure about it.

4.4.1 The problem

Here is a broad way to look at the laws of nature: we can think of them as governing or as non-governing; or as formulated in chapter 2, as allowing for potency or not. According to the governing conception laws impose constraints on what happens in the world. So, if laws have universal extent they govern not only non-actional events but also actional ones. If determinism happens to be true, then it is hard

to see how agents may have weak abilities. After all, events - including actional events - have no effects whatsoever on what laws of nature are. On the other hand, on the non-governing conception of laws, such as the Humean one, laws impose no constraints on what happens.

> If laws of nature are simply regularities, then the violation of a law (that is, the violation, in some nearby possible world, of an actual law of nature) isn't such a big deal. We happen to live in a world that is such that one set of regularities obtain; different sets of regularities obtain at other possible worlds. These are brute facts about the worlds in question, that are not to be explained by reference to some extra thing upon which those regularities depend. So there is no reason to think that the fact that my raising my hand would require a violation of the actual laws compromises my ability to do it. (Beebee 2003: 274)

Thus while (W) seems true according to the non-governing conception, it seems false according to the governing one.

With respect to (S), it is just the other way round. Since the governing conception rejects (W) it makes all the sense for its proponents to accept (S); agents do not have weak abilities, let alone strong ones.

> But from a Humean perspective it is very hard to see what reason there could be to believe it. If miracles are relatively cheap when it comes to events that are not actions of mine, or effects of actions of mine—as they must be if we are to hold that deterministic agents have weak abilities—why should they not be equally cheap when it comes to events that are actions of mine? (Beebee 2003: 274)

There is certainly a point here. Given the broad way to look at laws, it seems that neither the governing view nor the non-governing one can make both (W) and (S) to sound plausible at the same time. If this is the case, then LMC will in fact strike us as an untenable view. The counterexample to LMC may be unpersuasive, but there is still a problem that needs to be addressed. Let us put the problem in this schematic form:

1. If there is no account of laws consistent with LMC, then LMC is untenable

2. There is no account of laws consistent with LMC

   Thus,

3. LMC is untenable

My aim is to show that the local miracle compatibilist may have a good reason to deny 2. I will argue that the Humean view of laws is consistent with LMC. In particular, I will show that Lewis' version of the best system account of laws (LBSA) is consistent with LMC. I do not claim, however, that LBSA *per se* entails LMC. Naturally, the local miracle compatibilist does not need to provide an explanation as to why LMC and LBSA are consistent given only a certain view of laws and Lewis' argument about how counterfactuals should be evaluated on the assumption that laws L are deterministic. This may not be enough. As a result, an option for the local miracle compatibilist is to appeal to other principles to show why LMC and LBSA are consistent. The strategy is this. In order to show that propositions P and Q are consistent, we can show that P is consistent with another proposition R, and that the conjunction of P and R entails Q, so that P is consistent with Q. I will argue that this strategy will allow the local miracle compatibilist to have a reply to Beebee's problem. This will also motivate the claim that we never have strong abilities.

Since LMC was already explained, let me briefly rehearse Lewis' LBSA, as well as the principle I appeal to.

4.5 LMC and contrastive explanation

Here is how I formulated Lewis' BSA in chapter 2. Consider a true deductive system in which the general claims that represent laws of nature appear as a set of true sentences that is deductively closed and whose non-logical vocabulary contains only predicates that express perfectly natural properties. The laws of nature will belong to all the axiom systems with a best combination of simplicity, strength and fit.

The principle that will help us showing the consistency of Lewis' BSA with LMC is the following:

- **(Contrastive explanation)** If agent s performs some action a and there is no contrastive explanation of why s performed a rather than not-a, then a is not freely performed.

This principle comes up in Mele's formulation of the luck argument against libertarianism about free will (2006) as (roughly) presented in chapter 1. The argument is one of the biggest challenges to libertarians and the main point is that undetermined actions that lack contrastive explanation are not free. But the notion of an undetermined action will not be relevant for the purposes of the argument I put forward, for reasons that will be clear soon. Moreover, I do not claim that the argument is cogent. My claim is that *if* the argument is cogent, then the local miracle compatibilist can make sense of the distinction between weak and strong abilities.

In the original formulation, Mele conceives a scenario in which a goddess, Diana, creates "agents in an indeterministic universe which whenever they freely perform an action of deciding to a, they could have freely performed some alternative action" (2006: 8). Notwithstanding,

> She worries that her design does not accommodate this. Her worry, more specifically, is that if the difference between the actual world, in which one of her agents judges it best to A straightaway and then, at t, decides accordingly, and any possible world with the same past up to t and the same laws of nature in which he makes an alternative decision while the judgement persists is just a matter of luck, then he does not freely make that decision in that possible world, W. Diana suspects that his making that alternative decision rather than deciding in accordance with his best judgement— that is, that difference between W and the actual world—is just a matter of bad luck or, more precisely, of worse luck in W for the agent than in the actual world. After all, because the worlds do not diverge before the agent decides, there is no difference in them to account for the difference in decisions. (Mele, 2006: 8)

As Franklin correctly notes, Mele "is not claiming that the agent's decision in either the actual world or w is just a matter of luck" (Franklin 2011: 22). Rather, Mele's claim is that it is the cross-world difference between the actual world and w that is a matter a luck: that is, that the agent decided to do a at the actual world rather than decided not to do a as he did at w. And if this is a matter of luck, then it is partly a matter of luck that the agent decided to do a rather than not-a. If this is the case, then the decision is not free. Since Mele is considering the case in which the action is undetermined, we may conclude that indeterminism is also incompatible with free will.

How this may be used to motivate LMC? By definition, the most similar world to the actual at which a divergence miracle e occurs is a world whose events do not

diverge from those of the actual world until e itself occurs. Thus, the cross-world difference between the actual world and the most similar world where a divergence miracle occurs is a matter of luck, as there is nothing that accounts for this difference between these worlds. Now if a divergence miracle turns out to be an actional event, then it will not be a free action. Let us discuss the argument in its schematic form as presented by Franklin (Franklin 2011).

1. If agent s performs an undetermined action a at t in w and there is some world $w*$ that shares the same laws and past up to t, but in which s performs not-a, then there is nothing that accounts for the difference between world w in which s performed a and $w*$ in which s performed not-a.

2. If nothing accounts for this difference, then it is partly a matter of luck that s performed a in w and partly a matter of luck that s performed not-a in $w*$.

3. If an action is partly a matter of luck, then the action is not free

   Thus,

4. If s performs an undetermined action a at t in w and there is some world $w*$ that shares the same laws and past up to t, but in which S performs not-a, then both a and not-a are not free.

This formulation will allow us to explain why divergence miracles cannot be free actions. But there are some problems with Franklin's formulation of the argument (in the context of this discussion) that require clarification.

The first problem is with the expression "to share the same laws and past up to t". On a Humean view this does not make much sense, at least if literally interpreted. Given the assumption of determinism, w and w' cannot share the same past and laws up to t; after all, if this were the case, w and w' would be identical.

Although w and w' do not share the same laws and past up to t, the intuitive idea to be captured seems clear: a pair of worlds w and w' may be alike up until a certain time when they then diverge (thus "sharing the same past") and before they diverge the events of w' will be compossible with the laws of w (thus something pretty much like "sharing the same laws"). We can better express this in terms

of the notion of *an initial segment of a possible world* (Plantinga 1974; Wierenga 2011). I propose to use the notion of an initial segment of a possible world, rather than the notion of "sharing the same laws and past up to t". Following Wierenga in his paper "Tilting at Molinism" (Wierenga 2011: 127), I will adopt the following to talk about initial segments.

(i) For any world w and time t, there is a state of affairs, $\Sigma_{(w,t)}$, which is an initial segment of w terminating at t, and which is included in w.

I shall use this convention to talk about initial segments to make the formulation of the argument shorter:

(ii) Worlds w and w' share an initial segment up until a time t if, and only if, $\Sigma_{(w,t)} = \Sigma_{(w',t)}$

Furthermore, no world has more than one initial segment terminating at a certain time:

(iii) If $\Sigma_{(w,t)}$ and $\Sigma'_{(w,t)}$ are initial segments, then $\Sigma_{(w,t)} = \Sigma'_{(w',t)}$.

If two worlds share an initial segment up to a certain time, they share all of their initial segments terminating at earlier times:

(iv) If $\Sigma_{(w,t)} = \Sigma_{(w',t)}$, then for every t' such that t' is earlier than t, $\Sigma_{(w,t')} = \Sigma_{(w',t')}$

Notice that worlds may share an initial segment up until a time t without diverging at t, for they can continue to share initial segments after that time.

Finally, if two worlds share an initial segment, then the same things exist in both worlds, at least up until the time at which they diverge.

(v) If $\Sigma_{(w,t)} = \Sigma_{(w',t)}$, then for every x, x exists before t in w if and only if x exists before t in w'.

This is how I shall understand the notion of a pair of worlds "sharing the same past". Now we need to explain the notion of "sharing the same laws".

Because (on the assumption of determinism) Humean laws are the best summary of all past, present and future events, w and w' cannot have the same true propositions if they are distinct[14]. Let us say that a proposition P is true in an initial segment $\Sigma_{(w,t)}$ only if it is not possible that $\Sigma_{(w,t)}$ obtain and P be false. Since w and w' do not have the same laws, L cannot be true both in $\Sigma_{(w,t)}$ and $\Sigma_{(w',t)}$. But although $\Sigma_{(w,t)}$ and $\Sigma_{(w',t)}$ do not have the same laws, it must be the case that if L is true in w, L is compossible with $\Sigma_{(w',t)}$. After all, w and w' are exactly alike up until t. What we can say instead is that the segments are still nomologically accessible, for the events of $\Sigma_{(w',t)}$ are compossible with the laws of w.

This explains why premise 1 does not need to be formulated in terms of an undetermined (actional) event, that is, on the assumption that laws L are indeterministic. According to a governing conception of laws (one that does not allow for potency), the events of a world happen on account of the laws of that world, so that if laws are deterministic in Lewis' sense it cannot be that two distinct worlds share initial segments. On a Humean view, on the other hand, this is possible. Since the things that L speaks about happen merely in accord with L, if a pair of worlds shares an initial segment up until a time t the worlds will be in accord with L up until t.

Now, suppose that worlds w and w' share an initial segment up until a time t, and suppose that someone performing an action leads to a divergence of worlds. In other words, suppose that agent s performs action a at t in w, but not at t in w', so that $\Sigma_{(w,t)} = \Sigma_{(w',t)}$ and w and w' diverge at t:

1. If s performs action a at t in w and there is some world w' such that $\Sigma_{(w',t)} = \Sigma_{(w,t)}$, but in which s performed not-a at t, then there is nothing up to t that accounts for the difference between w in which s performed a at t and w' in which s performed not-a at t.

2. If nothing up to t accounts for this difference, then it is partly a matter of luck that s performed a in w and partly a matter of luck that s performed not-a in w'.

[14] An alternative though is to combine a Humean conception of laws with the Growing Block Theory, where future events do not exist. See Backmann (2016).

3. If an action is partly a matter of luck, then the action is not free

   Thus,

4. If s performs action a at t in w and there is some world w' such that $\Sigma_{(w',t)} = \Sigma_{(w,t)}$, but in which s performed not-a at t, then both a and not-a are not free.

Suppose that a is Beebee's decision to raise her hand. In the actual world a does not occur at t, since she does not decide to raise her hand, but there is some possible world w' where a occurs and is the divergence miracle relative to the actual, so that $\Sigma_{(w,t)} = \Sigma_{(w',t)}$. If this is the case, given premise 2, there is nothing that accounts for the difference between the actual world in which Beebee did not decide to raise her hand and w' in which she did decide to raise her hand. Thus, clearly, both actions are partly a matter of luck. So here we have the problem. For it cannot be that a free action is partly a matter of luck.

This, of course, can be generalised to every action that is supposed to be a divergence miracle. If a divergence miracle e is an action at world w relative to the actual world, then the events of w do not diverge from the actual until e itself occurs. And there is nothing that accounts for this cross-world difference. Therefore, there cannot be free actions that are divergence miracles.

My suggestion then is to formulate LMC as follows:

(W) Some agents have the ability to perform a *free* action a such that, if someone were to do a, there would be an event e such that necessarily if e occurs L is false.

(S) Agents do not have the ability to perform a *free* action a such that, if someone were to do a, there would be an event e such that necessarily if e occurs L is false and e is identical to a.

The plausibility of such a view depends on the formulation of the argument being sound. But why should someone accept its premises?

Premise (1). If two worlds are exactly alike up to t, then what would account for the difference between them at t? Up to t there is certainly nothing to account for that. We cannot appeal, for example, to prior mental states that gives reasons

explanations since the worlds share an initial segment up until t and so agree on the occurent facts up until t.

Perhaps one might follow Mele's suggestion that "in a deterministic world, with the possible exception of chaotic events, all events are explicable, at least in terms of laws and antecedent conditions" (Mele 2000: 98). This can be reformulated as an objection in the following way. Given the assumption that laws L are deterministic at w, then L cannot be true at w', so that the laws will explain the difference.

I believe Mele's point is correct only if we are willing to assume a governing view of laws in which the events happen on account of L. But this is not the case on the Humean view. The expression "to account for the difference" can be understood in terms of metaphysical grounding. Humean laws cannot ground the difference between two identical initial segments simply because the laws do not ground the mosaic of occurrent features. This is not to say that Humean laws cannot provide genuine explanations. This is compatible with Humean laws scientifically explaining in the sense that one may use them to make predictions. But since they do not ground the difference at t between the worlds, they do not account for the difference, they do not *metaphysically explain* the difference.

One might object as follows. It is fair enough that Humean laws do not ground occurent events. Still, they do allow us to make predictions. And if this is the case, it cannot be that the action - even if a divergence miracle - is lucky. The problem with the above objection is to suppose that being able to predict an event is sufficient for that event not to be lucky. I am not sure about this. A person can be lucky because she was born in a certain country, or in a certain family, even though - assuming laws to be deterministic - this can be quite predictable.

This leads us to premise (2). In order to defend premise (2) I need to clarify what is meant by "luck". In the original formulation of the argument, Mele provides us with a stipulative definition of "luck" precisely in terms of the unavailability of contrastive explanation. It is clear that if one defines "luck" in that way the premises will come out as true. As Franklin has correctly noted, though, this strategy is fallacious (Franklin 2011: 221). I am persuaded by Franklin's argument and I follow his suggestion that the best way out for proponents of the luck argument is to "leave

the notion of luck unanalyzed, and argue instead that since nothing accounts for the difference between worlds, each action is partly a matter of luck — where luck here is understood in an intuitive sense" (Franklin 2011: 222). What is the intuitive sense of luck then?

One strategy for the local miracle compatibilist is to argue that "luck" means lack of control in the following sense: "the more an action is subject to luck, the less it is under our control, and the more an action is under our control, the less it is subject to luck" (Franklin 2011: 200). As Markus E. Schlosser (2014) points out there is a connection between explanation and control, and I take it to be open to the local miracle compatibilist to track control in terms of explanation. For example, Schlosser suggests that according to the event-causal theory "control consists in non-deviant causation by mental states that provide [...] reason explanations. This is not to say that all explanations track control but rather that if the right type of explanation is available, then it will track control tightly" (Schlosser 2014: 382). If this is right, then it is open to the local miracle compatibilist to track control in terms of explanation.

Now, if it is uncontroversial that Humean laws *scientifically explain*, why cannot we just say that they account for the difference? And so the divergence will not be a matter of luck. Leaving aside the question of whether Humean laws provide *genuine explanation*, it is not clear whether Humean laws may provide a sort of explanation that tightly track control. The problem is that such a sort of explanation will have to appeal to *later* events. What we need - it seems - is something that accounts for the difference *up until* t. Imagine a scenario where w and w' come to an end at t even though they share initial segments up until t. If we say that L explains the difference, we are just saying that what explains the difference is s performing a at t, for laws L are different at worlds w and $w*$ just because s performed a at w but not at w'. But this cannot be right. After all libertarians could just deny the first premise of the luck argument by saying that the performance of the action itself explains the difference[15].

[15]Well, there is such a reply to the argument (see van Miltenburg & Ometto forthcoming), but it is quite controversial (Mele 2014; Clarke forthcoming). I will come back to this point in chapter

The link between divergence miracles and lack of control can be put in terms of the event-causal theory. Suppose control consists of (non-deviant) causation by mental states that provide reason explanations. Imagine two possible worlds where I have both a reason r_1 to perform a at t and a reason r_2 to perform not-a at t, and suppose that I do a. We may ask: why did I do a? The answer will just be that r_1 played a causal role in bringing about my action, whereas r_2 did not, and this explains the difference. However, if r_1 played the causal role, then r_1 is an occurrent mental state, that is, it is a "going on" or "happening" in my consciousness, as opposed to a dispositional mental state[16]. However, when we consider the world at which I decided not-a, r_2 will have to be the occurrent mental state that does the causing, and r_2 will have to occur earlier than t, the time at which I decided not-a. This contradicts the assumption that the worlds are exactly alike up to t, as at w_1 r_1 was the occurent mental state.

Premise (3). If an action is partly a matter of luck, then the action is not free. This is the least problematic premise of the argument. Since by "luck" it is meant "lack of control" it is clear that not having control over an action implies that the performance of that action is not free.

Here is a worry, though. If the luck argument is sound, even my decision to raise my hand at t in w will not be free. This looks implausible.

I agree that this seems implausible. Nevertheless, this does not mean that all actions are not free according to LMC: only those actions that lack a contrastive explanation will not be free. What LMC might require is that the cause of a free action is not an earlier free action, but an earlier non-free action. If it happens that a divergence miracle is the activation of the reasons to do otherwise, then this will explain the cross-world difference. To go back to that earlier case where I have both a reason r_1 to perform a at t and a reason r_2 to perform not-a at t, what will explain the difference is that r_1 plays the causal role at one world and r_2 at another world.

7.

[16] I follow Alvin Goldman (1970) in saying that dispositional wants and beliefs do not by themselves cause acts; they "can affect action only by becoming activated, that is, by being manifested in occurrent wants and beliefs" (Goldman 1970: 88).

In any case, my free action is caused by a non-free action (say, the activation of the reasons).

This completes the presentation of the argument. The upshot is that this formulation of the luck argument, if cogent, may be an advantage for the local miracle compatibilist, but a problem for compatibilists who endorse a Humean view of laws, but deny Lewis' distinction. In articulating "Humean compatibilism", Beebee and Mele (2002) had already raised this worry about luck:

> [A] Humean compatibilist holds that an agent, Barney, could have done otherwise than decide to steal his neighbour's cake if and only if there is a possible world that... is exactly like the actual in every detail up to the moment at which he decided to steal her cake, and, in it, he does something other than decide to steal it.
>
> A worry about luck leaps out. What accounts for its being the case that although, in the actual world, Barney decides at t to steal the cake, in another 'world that ... is exactly like the actual world in every detail until t, Barney decides at t to go bowling instead? Apparently, there is nothing about the powers, capacities, states of mind, moral character, practical reasoning and the like of Barney the cake stealer and Barney the bowler that explains the difference in decision, given that the two worlds are exactly the same until t. So the difference seems to be a matter of luck. A Humean compatibilist may attempt to account for the difference in decision by appealing to a difference in laws; but if the pertinent laws themselves hinge on Barney's decisions, this smacks of unacceptable bootstrapping. (Beebee & Mele 2002: 221)

There is an important difference between the view above and local miracle compatibilism as articulated here. The difference can be explained if we consider the following principles:

(Alternative possibilities) If agent s freely performs some action a at t in w, then there is another possible world $w*$ that is exactly like w up until t where s refrains from performing a at t.

(Contrastive explanation) If nothing accounts for the difference between world w in which s performs a at t and $w*$ in which s refrains from performing a at t, then a is not free.

If there is free will, clearly, we cannot accept both principles. If one accepts the principle of contrastive explanation, then one has to deny the principle of alternative possibilities.

Now, of course, the plausibility of local miracle compatibilism depends on the truth of the principle of contrastive explanation. I believe that a correct response to this problem will depend on what control is, which is something I am not dealing with in the thesis. But either the explanatory formulation of the luck argument is sound or not. If it is sound, then LMC is not an untenable view. If the argument is not sound, then one can abandon the claim that agents do not have strong abilities and adopt a Humean compatibilist view as developed by Beebee and Mele (after all, they will not have to worry about luck anymore). In any case, I think that at least some Humean compatibilist view motivates the rejection of premise 2 in the dilemma presented in chapter 1.

4.6 Conclusion

If this formulation of the luck argument is sound, then *free* actions cannot be divergence miracles. So I suggest to formulate LMC as follows:

(W) Some agents have the ability to perform a *free* action a such that, if someone were to do a, there would be an event e such that necessarily if e occurs L is false.

(S) Agents do not have the ability to perform a *free* action a such that, if someone were to do a, there would be an event e such that necessarily if e occurs L is false and e is identical to a.

I have presented the luck argument to show that (W) and (S) are consistent with the assumption that the actual laws are deterministic and universal. All that the local miracle compatibilist needs to do is to accept the connection between explanation and control, such as the idea of tracking control in terms of the causal theory of action. If some actions are divergence miracles then they will lack any type of explanation that will track control.

4.7 Addendum: reply to Westphal

Westphal has recently objected to FFA by pointing out a series of difficulties in van Inwagen's argument, and the most serious one is - according to him - a modal fallacy presented in support of the second premise. My claim is that Westphal's objection fails. First, I will briefly put forward FFA and Westphal's objection. After that, I will go on to argue that there is no modal fallacy presented in support of its second premise. The argument presented in *An essay on free will* is slightly different from the original one (that I discussed before) because van Inwagen changed the formulation of premise 2:

1. The truth of determinism entails that the conjunction of P_0 and L entails P.
2. It is not possible that j have raised his hand at t and P be true
3. If 2 is true, then if j could have raised his hand at t, j could have rendered P false.
4. If j could have rendered P false, and if the conjunction of P_0 and L entails P, then j could have rendered the conjunction of P_0 and L false.
5. If j could have rendered the conjunction of P_0 and L false, then j could have rendered L false.
6. j could not have rendered L false.

   Therefore,

7. If determinism is true, j could not have raised his hand at t.

Westphal objects to van Inwagen's argument for the second premise, calling it invalid. Here is what van Inwagen gives in support of (2):

> "The symbol 'P' is our name for the proposition that expresses the state the world was in fact in at t, a time at which j's hand was not raised. It is therefore impossible for P to be true if j's hand was raised at t, or indeed if things were in any way different at t from the way they actually were." (van Inwagen, 1983: 70)

van Inwagen writes that it is impossible for P to be true *if* [my emphasis] j's hand was not raised at t. Following Westphal I will use the italic capital letter J for the proposition that the judge did raise his hand at t, and the italic capital letter P for the proposition denoted by "P".

4.7.1 Westphal's objection

Westphal argues in his paper that either plausible reading of van Inwagen in the argument for (2),

(K) $\sim \Diamond (J \supset P)$

or

(C) $J \supset \sim \Diamond P$,

is false.

Consider (K). If (K) is true, then it follows that $\Box J$ and $\Box \sim P$. However, it cannot be that the conjunction is true. Since j doesn't raise his hand, $\sim J$ is true, and so it follows (assuming a modal system at least as strong as T) that $\Diamond \sim J$, which contradicts $\Box J$.

Now consider (C). Westphal argues that the argument for (C) is a modal fallacy. If J is true, it only follows that $\sim P$. It doesn't follow that P *cannot* be true.

> What is wrong with what van Inwagen actually writes is to be found in the proposition that if p is true, then p cannot be true. It is one thing to say that (i) it is a necessary truth that, if Tp, then $T\sim\sim p$. It is quite another thing to say (ii) that if Tp, then it is a necessary truth that $F\sim p$. For (i) is true and (ii) is false. And so most certainly (ii) does not follow from (i). (Westphal 2012: 38)

Westphal thinks that van Inwagen genuinely meant to assert $J \supset \sim \Diamond P$. I do not agree with him. Instead I contend that there is a more charitable way to understand the argument for (2).

First, van Inwagen could just have said that (2) is conceptually true. What the expression "it is therefore impossible for P to be true if j's hand was not raised" means in English is that it is impossible for P to be true *and* j's hand to be raised at t. If Westphal's point is that van Inwagen used the wrong connective, then one

might just say that this is no modal fallacy but a careless way to state a necessary truth. In that case, he should have written:

(2*) "It is impossible for P to be true *and* [my emphasis] j's hand to be raised at t."

That is, (2*) should be read as:

(2*) $\sim \Diamond (P \wedge J)$

This is a necessary truth that falls out of the definition of "P". P is a proposition that describes the whole state of the world at t, one that includes j not raising his hand. That is, necessarily, if P is true, then j does not raise his hand. Contrapositively, necessarily, if j raises his hand, P is not true, which is equivalent to (2*). Thus, it is impossible for P to be true and j's hand to be raised.

This is the correct argument for the second premise of van Inwagen's "First formal argument", and it is entirely unproblematic. Consequently, Westphal's arguments against van Inwagen's presumed motivations for $J \supset \sim \Diamond P$ are beside the point and leave the "First Formal Argument" unscathed.

(2*) allows one to avoid the charge of modal fallacy, and it is compatible with Westphal's understanding of premise (2). However, the premise as stated in the original argument (in 1975) is not a conjunction, but a conditional (van Inwagen 1975: 191):

(2) If j had raised his hand at t, P would be false.

Since (2*) is true, it follows that, necessarily, if j raises his hand at t, then P is false, that is, $\Box(J \supset \sim P)$. So it follows that if j had raised his hand at t, P would be false, simply because the strict conditional implies the counterfactual one. Thus, premise (2) as originally formulated is true.

Moreover, in the first sentence of the paper Westphal claims: "I believe that the argument given by Peter van Inwagen for the second premiss in his 'First Formal Argument' (van Inwagen 1983: 70) is invalid, and that accordingly the entire 'First Formal Argument' is unsound" (Westphal 2012: 36). Even if, counterpossibly, the argument for the second premise was invalid, which it can't be since the conclusion

is a necessary truth, it would not follow that the second premise is false, so it would not follow that the argument is not sound.

Chapter 5

The dispositional conception of laws and agent causation

In the previous chapter I argued that at least some Humean compatibilist view motivates the rejection of premise 2 in the dilemma presented in chapter 1[1]. In this chapter I explore the consequences of adopting (i) a completely anti-Humean account of laws and (ii) O'Connor's metaphysical assumptions of agent causation. With respect to (i), I show that it justifies the premise of the consequence argument that the laws of nature are not up to us. I will argue that the premise is necessarily true given the assumption that laws are DE-deterministic (as defined in chapter 2) and universal in extent, and the counterfactual sufficiency interpretations of "no-choice about" (as discussed in chapter 3). With respect to (ii), I explore the view that free will requires the laws of nature to be limited in extent; that is, the distinctive characteristic of this view is that it requires laws to be limited in extent (because they do not apply to agents), rather than indeterministic.

Here is how the chapter is structured. In section 5.1 I briefly rehearse the consequence argument and present Warfield's modal fallacy objection. After that, I argue that the dispositionalist view of laws allows the incompatibilist to reply to

[1]As pretty much every philosophical view, there are of course some problems with this sort of position, and they include the objections to the Humean view of laws (cf. Demarest 2017; Callender and Cohen 2009), and Lewis' argument for the existence of miracles, which has been recently challenged (Dorr 2016; Goodman 2015).

Brian Cutter's (2017) objection to the consequence argument. In sections 5.3 and 5.4, I discuss the objection that dispositional essentialism is incompatible with the Lewis-Stalnaker semantics for counterfactuals, one I have adopted to interpret the common phrases of the consequence argument such as "to have the power to make P false", "to make it the case that $\sim P$, "to have a choice about whether P", etc. In order to reply to this problem I appeal to Handfield's solution (2001, 2004) in terms of the notion of space-invaders. After that, in section 5.5, I consider the theory of agent causation and draw some parallels between space-invaders and the view that free will requires the laws of nature to be limited in extent.

5.1 The charge of modal fallacy

Very briefly, let us consider the "counterfactual sufficiency" definition of "**N**". Again, where x ranges over agents and α ranges over all possible past, present and future action types:

Definition 5.1.1. *$\boldsymbol{N}\phi$ if and only if $\phi \wedge \sim\exists x \exists \alpha[Can(x,\alpha) \wedge (Does(x,\alpha) \boxdot\!\!\rightarrow \sim\phi)]$*

The consequence argument may be formulated as follows on Stalnaker's theory (where the box is to be understood in terms of metaphysical necessity):

(α) $\Box_M \phi \vdash \mathbf{N}\phi$

(β) $\mathbf{N}\phi, \mathbf{N}(\phi \rightarrow \psi) \vdash_{Stalnaker} \mathbf{N}\psi$

Suppose that laws are deterministic and universal. Here is the argument once again:

1	$\Box_M((\mathcal{L} \wedge P_{t0}(L,P)) \supset P)$	DE-deterministic
2	$\Box_M(\mathcal{L} \rightarrow (P_{t0}(L,P) \supset P))$	Modal logic, 1
3	$\mathbf{N}(\mathcal{L} \rightarrow (P_{t0}(L,P) \supset P))$	α, 2
4	$\mathbf{N}\mathcal{L}$	premise
5	$\mathbf{N}P_{t0}(L,P)$	premise
6	$\mathbf{N}(P_{t0}(L,P) \supset P)$	β, 3, 4
7	$\mathbf{N}P$	β, 5, 6

Warfield (2000) has objected to it by pointing out that, as long as it is an argument for incompatibilism, it needs to be an argument for the claim that determinism *strictly implies* $\mathbf{N}P$ (or $\mathbf{M}P$ if we formulate it on Lewis' theory of counterfactuals). Let us consider the second proposition of the dilemma presented in chapter 1:

(2) If determinism is true, then $\mathbf{N}P$.

(2) is supposed to be *incompatibilism*, the view that determinism (in the standard sense) and free will are incompatible. However, I have not said much about the meaning of "if-then" in (2). Warfield's point is that if by "if-then" we mean the material conditional, then (2) is not a thesis that may be properly called "incompatibilism". (2) understood in terms of the material conditional does not establish the *incompatibility* of free will and determinism. In this sense, one might be an incompatibilist just because she thinks determinism is not true. But this cannot be right.

Now, if (2) is a thesis that may be properly called "incompatibilism", then a better way to understand it is in terms of the strict implication. That is, rather than being an argument for

(Weak) If determinism is true, then $\mathbf{N}P$,

where the "if-then" of (Weak) does not have the force of the strict implication, it should be an argument for

(INC) Necessarily, if determinism is true, then $\mathbf{N}P$.

(Here I use "determinism" in the sense that laws are universal and deterministic). As Warfield correctly notices, (INC) does not follow from the premises if they are contingently true:

> Most incompatibilists, to be precise, seem unaware that in order to get the incompatibilist conclusion that determinism and freedom are strictly incompatible (that no deterministic world is a world with freedom), their conditional proofs must not introduce or in any way appeal to premises that are merely contingently true in between the assumption of determinism and the step at which the "no freedom" conclusion is reached. (Warfield 2000: 169)

This generates a problem if one is an incompatibilist and thinks - perhaps by adopting the Armstrong/Tooley/Dretske view of laws - that laws are contingent. After all, "**N**" is a factive operator. If $\Box\mathbf{N}\mathcal{L}$ is true, then it will follow that $\Box\mathcal{L}$. Because $\mathcal{L}$ is contingently true, it follows that $\Box\mathbf{N}\mathcal{L}$ is not true. Thus, the consequence argument in its standard formulation is an *ignoratio elenchi*: at best, it merely establishes (Weak), not (INC).

Hence, if one adopts the Armstrong/Tooley/Dretske view of laws, one cannot justify the premise that $\Box\mathbf{N}\mathcal{L}$. What someone who adopts this view can do is to appeal to the following principle:

Necessary Fixity of Laws (*NFL*): for every world w, it's true in w that $\mathbf{N}L_w$ (where L_w is a conjunction of all the laws of nature that prevail at w).

But why should one think that *NFL* is true? Cutter (2017) has recently challenged *NFL*. His argument appeals, as he says, "to creatures in the far reaches of modal space, creatures whom" he calls *miracle workers* (Cutter 2017: 283). First, Cutter's idea is that the possibility of miracle workers entails the falsity of *NFL*[2]. He defines a miracle worker as follows: s is a miracle worker in w if, and only if, s has, in w, the ability to do something such that, if he did it, the laws that prevail in w would be violated. (What he means by a miracle worker is tantamount to Lewis' weak thesis presented in the last chapter). And he also introduces the notion of a *coy* miracle worker: a coy miracle worker is a miracle worker who never exercises the ability to violate the laws of nature. What premise 4 in the consequence argument tells us is that no one is a miracle worker. Cutter agrees that this premise is actually true, but disagrees that it is necessarily true.

> Although I find it implausible to suppose that actual human beings are miracle workers [...] I do not have any trouble with the claim that it's possible for there to be miracle workers. After all, there are all sorts of bizarre things out in the far reaches of modal space – seven-headed monsters, golden mountains, and talking donkeys, for instance. And if it's possible that there should be miracle workers, then surely it's possible that there should be coy miracle workers. (Cutter 2017: 283)

[2]Actually, Cutter has argued for a more shocking conclusion: that the consequence argument is not even an argument for incompatibilism.

Cutter does not claim that it is obvious that miracle workers are possible. He claims that "we do not have good positive reason to think they are not possible" (Cutter 2017: 284). For example, he tells us that the possibility of miracle workers is compatible with the view that a proposition P is a law of nature at w only if P is true in w. What follows from this view is that:

- For each world w and each agent s in w, s does not act, in w, in such a way that the laws that prevail in w are violated/false.

The only thing that follows from it is that, necessarily, every miracle worker is a coy miracle worker. Cutter also points out that this does not mean, naturally, that "coy miracle workers are such that it's impossible for them to exercise their abilities. It just means that if they were to exercise their abilities, the proposition they would thereby falsify would, in that event, not qualify as a law of nature" (Cutter 2017: 284).

Are miracle workers impossible according to the Armstrong/Tooley/Dretske view of laws? I am not entirely sure. Perhaps what the governing view justifies is the claim that actual human beings are not miracle workers, but not the claim that miracle workers are impossible. Perhaps not. Although I agree with Beebee that the governing view seems more in line with incompatibilism, I do not know how to show the impossibility of miracle workers according to the Armstrong/Tooley/Dretske view of laws. Even so, I think I can show that miracle workers are impossible according to dispositional essentialism.

5.2 Dispositional essentialism and laws

As we have seen in chapter 2, dispositional essentialism is the view according to which at least some natural, fundamental properties are essentially dispositional (although Bird himself adopts the stronger view that all, natural, fundamental properties are essentialy dispositional). Dispositional essentialists follow Sydney Shoemaker's suggestion (1980) that properties are "individuated" by their causal

powers[3]. Chris Swoyer (1982) developed it as view about laws of nature, and the connection between essentially dispositional properties and laws has also been promoted by Ellis (2001, 2002), Bostock (2001), Kistler (2002) and Bird (2005, 2007). Here, I will mainly focus on Bird's dispositional account of laws, but my argument depends only on the view that the laws of nature are metaphysically necessary. My choice for presenting Bird's dispositional account of laws is because it will make the connection between the laws of nature and free will much more precise and simpler to be discussed.

On Bird's account, the truth of generalisations can be explained on the basis of the dispositional essence of a property. A consequence of this view is that the laws of nature are metaphysically necessary. If this is so, it will follow that no one has or ever had any choice about whether $\mathcal{L}$ is true. In other words, I will argue that if laws are DE-deterministic as defined in chapter 2 and universal, miracle workers are impossible.

I start by briefly recapitulating Bird's dispositional account of laws presented in chapter 2. Here is how the dispositionalist may account for laws of nature. Start first with the conditional analysis of dispositions (CA for short). Where D is a dispositional property, $S(D)$ is a stimulus property appropriate to it and $M(D)$ is its manifestation property, (CA) may be symbolised as follows:

$$\text{(CA)}\ \ Dx \leftrightarrow (S(D)x \boxdot\!\!\rightarrow M(D)x)$$

As Bird points out, (CA) does not merely provide an analysis of the concept D; instead, it characterises the nature of the property D. Thus, as Bird says, (CA) is to be read as being metaphysically necessary.

$$\text{(CA}_{\Box}\text{)}\ \ \Box_M(Dx \leftrightarrow (S(D)x \boxdot\!\!\rightarrow M(D)x))$$

Second, dispositional essentialists endorse the view that at least some fundamental properties are essentially dispositional. To say that a property P is essentially dispositional is to say that, necessarily (in the metaphysical sense), to instantiate P

[3]Dispositional essentialists include, among others, Harré and Madden (1975), Ellis and Lierse (1994), Ellis (2001, 2002), Molnar (2003), and Bird (2005, 2007).

is to possess a disposition $D(P)$ to yield the appropriate manifestation in response to an appropriate stimulus:

(DE_p) $\Box\ (Px \rightarrow D(P)x)$

From $(\text{CA}_\Box)$ and (DE_p) we have: $\Box_M(Px \rightarrow (Sx \ \Box\!\!\rightarrow Mx))$. Here is how the truth of generalisations can be explained on the basis of the dispositional essence of a property.

1	$\Box_M(Px \rightarrow (Sx \ \Box\!\!\rightarrow Mx))$	
2	$\quad Px \rightarrow (Sx \ \Box\!\!\rightarrow Mx))$	
3	$\quad\quad Px \wedge Sx$	
4	$\quad\quad Px$	$\wedge$E, 3
5	$\quad\quad Sx$	$\wedge$E, 3
6	$\quad\quad Sx \ \Box\!\!\rightarrow Mx$	$\Rightarrow$E, 2, 4
7	$\quad\quad Mx$	$E\ \Box\!\!\rightarrow$, 5, 6
8	$\quad (Px \wedge Sx) \rightarrow Mx$	$\Rightarrow$I, 3–7
9	$\quad \forall x((Px \wedge Sx) \rightarrow Mx)$	$\forall$I, 8
10	$\Box_M(\forall x((Px \wedge Sx) \rightarrow Mx))$	$\Box$I, 1, 2–9

I think I can now reply to the modal fallacy challenge (at least with respect to premise 4).

Premise 4 is true if and only if $\mathcal{L} \wedge {\sim}\exists x \exists \alpha[Can(x, \alpha) \wedge (Does(x, \alpha) \ \Box\!\!\rightarrow {\sim}\mathcal{L})]$. Now, because $\mathcal{L}$ is necessarily true, there is no situation in which the counterfactual "$Does(x, \alpha) \ \Box\!\!\rightarrow {\sim}\mathcal{L}$" is true. Or to put it in another way, consider rule (α). Given this rule we can derive $\mathbf{N}\mathcal{L}$. Perhaps one may wonder whether rule (α) is correct (see van Miltenburg and Ometto forthcoming). This should not worry us. It can be proved on Stalnaker's system that the inference is legitimate.

(S) $\phi \ \Box\!\!\rightarrow \psi,\ \Diamond\phi \vdash_{Stalnaker} {\sim}(\phi \ \Box\!\!\rightarrow {\sim}\psi)$

1	$\Box_M \mathcal{L}$	
2	$\mathcal{L}$	$E\Box$, 1
3	$\quad Can(s,a) \wedge (Does(s,a) \; \Box\!\!\rightarrow \sim\mathcal{L})$	
4	$\quad Does(s,a) \; \Box\!\!\rightarrow \sim\mathcal{L}$	$\wedge$E, 3
5	$\quad\quad Does(s,a)$	
6	$\quad\quad \mathcal{L}$	$E\Box$, 1
7	$\quad Does(s,a) \; \Box\!\!\rightarrow \mathcal{L}$	$I \; \Box\!\!\rightarrow$, 5–6
8	$\quad Can(s,a)$	$\wedge$E, 3
9	$\quad Can(s,a) \rightarrow \Diamond Does(s,a)$	
10	$\quad \Diamond Does(s,a)$	$\Rightarrow$E, 8, 9
11	$\quad \sim(Does(s,a) \; \Box\!\!\rightarrow \mathcal{L})$	S, 4, 10
12	$\quad \bot$	
13	$\sim(Can(s,a) \wedge (Does(s,a) \; \Box\!\!\rightarrow \sim L))$	$\neg$I, 3–13
14	$\forall x \sim(Can(x,a) \wedge (Does(x,a) \; \Box\!\!\rightarrow \sim\mathcal{L}))$	$\forall$I, 13
15	$\forall x \forall \alpha \sim(Can(x,\alpha) \wedge (Does(x,\alpha) \; \Box\!\!\rightarrow \sim\mathcal{L}))$	$\forall$I, 14
16	$\sim\exists x \exists \alpha [Can(x,\alpha) \wedge (Does(x,\alpha) \; \Box\!\!\rightarrow \sim\mathcal{L})]$	Logic, 15
17	$\mathcal{L} \wedge \sim\exists x \exists \alpha [Can(x,\alpha) \wedge (Does(x,\alpha) \; \Box\!\!\rightarrow \sim\mathcal{L})$	$\wedge$I, 2, 16

Since 17 is equivalent to $\mathbf{N}\mathcal{L}$, given definition (5.1.1), the dispositionalist account of laws allows us to show why no one has or ever had any choice about whether $\mathcal{L}$ is true. We can get the same conclusion if we interpret "no-choice about" in terms of Lewis' might-counterfactual.

Definition 5.2.1. *$\boldsymbol{M}\phi$ if and only if* $\phi \wedge \sim\exists x \exists \alpha [Can(x,\alpha) \wedge Does(x,\alpha) \; \Diamond\!\!\rightarrow \sim\phi)]$

The following proof can be given on Lewis' system:

(L) $\phi \; \Diamond\!\!\rightarrow \psi \dashv\vdash_{Lewis} \sim(\phi \; \Box\!\!\rightarrow \sim\psi)$

1	$\Box_M \mathcal{L}$	
2	L	$E\Box$, 1
3	$\quad Can(s,a) \wedge (Does(s,a) \diamond\!\!\rightarrow \sim\mathcal{L})$	
4	$\quad Does(s,a) \diamond\!\!\rightarrow \sim\mathcal{L}$	$\wedge$E, 3
5	$\quad\quad Does(s,a)$	
6	$\quad\quad \mathcal{L}$	$E\Box$, 1
7	$\quad Does(s,a) \Box\!\!\rightarrow \mathcal{L}$	$\Rightarrow$I, 5–6
8	$\quad Does(s,a) \diamond\!\!\rightarrow \sim\mathcal{L}$	$\wedge$E, 3
9	$\quad \sim(Does(s,a) \Box\!\!\rightarrow \mathcal{L})$	L, 8
10	$\quad \bot$	
11	$\sim(Can(s,a) \wedge (Does(s,a) \diamond\!\!\rightarrow \sim L))$	$\neg$I, 3–10
12	$\forall x \sim(Can(x,a) \wedge (Does(x,a) \diamond\!\!\rightarrow \sim L))$	$\forall$I, 11
13	$\forall x \forall \alpha \sim(Can(x,\alpha) \wedge (Does(x,\alpha) \diamond\!\!\rightarrow \sim\mathcal{L}))$	$\forall$I, 12
14	$\sim\exists x \exists \alpha [Can(x,\alpha) \wedge (Does(x,\alpha) \diamond\!\!\rightarrow \sim\mathcal{L})]$	Logic, 13
15	$\mathcal{L} \wedge \sim\exists x \exists \alpha [Can(x,\alpha) \wedge (Does(x,\alpha) \diamond\!\!\rightarrow \sim\mathcal{L})$	$\wedge$I, 2, 14

The upshot is that, contrary to Humeans, dispositional essentialists should be committed to the truth of the premise that the laws of nature are not up to us. But remember that the main challenge is to argue that the premise is necessarily true. Now we can also show how this view supports the strong version which Warfield says incompatibilists should argue for. The charge of modal fallacy can be avoided by appealing to the S4 axiom: $\Box\phi \rightarrow \Box\,\Box\,\phi$. And dispositionalists may proceed as before in showing how it is necessary that the laws of nature are not up to anyone. That is,

1	$\Box_M \mathcal{L}$	
2	$\Box_M\, \Box_M\, \mathcal{L}$	S4, 1
3	$\quad \Box_M \mathcal{L}$	
4	$\quad \mathbf{N}\mathcal{L}$	α, 3
5	$\Box_M \mathbf{N}\mathcal{L}$	$E\Box$, 2, 3–4

Thus, dispositional essentialism allows us not only to show why premise 4 is true but also that it is necessarily so. Moreover, there is no need to appeal to a principle such as the *necessary fixity of laws*. Thus, we have a reason to think that miracle workers are impossible: the dispositionalist conception of laws. According to this view, it follows that $\Box_M \mathbf{N}\mathcal{L}$, which is just the denial of the claim that miracle workers are possible.

(A final point: if this is correct, then premise 4 is true. This is not enough to establish incompatibilism, though. There is still the question of whether $\mathbf{N}P_{t0}(L, P)$ is necessarily true. And Campbell has argued (2007) that that premise, even if it is true, is not necessarily true. If the main point is to argue for incompatibilism given the dispositional account of laws, Campbell's objection needs to be discussed. This will be done in the next chapter.)

5.3 An objection to dispositional essentialism

One worry that someone might have is that I have used the Lewis-Stalnaker semantics in order to show premise 4 is true according to DE. However, as we have seen in the last chapter, Lewis presented an argument to tell us how we should evaluate true counterfactuals on the assumption that laws L are deterministic and universal. This argument may be presented as an objection to dispositional essentialism. Remember that I have defined DE-deterministic laws as follows:

Definition 5.3.1. *Laws L are DE-deterministic if and only if for any P that L covers and any $P_{t0}(L, P)$ and any metaphysically possible worlds w, w' in which $\mathcal{L}$, if w and w' agree on $P_{t0}(L, P)$ they agree on whether P obtains.*

How should we evaluate counterfactuals on the assumption that laws L are DE-deterministic and universal? It seems Lewis' argument could be put forward for this definition. Following work with Cartwright I will put forward the argument in the following way.

Let P stand for the proposition that I raise my hand, suppose that P is false at the actual world w_0. Assuming DE-laws are deterministic and universal, we ask: is $\mathcal{L}$ true in all the most similar worlds at which P is true? If we consider Lewis'

argument as presented in chapter 4, the answer is "no". Lewis' suggestion is to regard as "most similar" those worlds where the past - but not the laws - are fixed.

Clearly, however, the above manoeuvre is not available to dispositional essentialists. If laws are metaphysically necessary, there cannot be any divergence miracles. And if this is the case, then Lewis' overall comparative similarity ranking will be at odds with dispositional essentialism[4].

> 'So much the worse for possible-worlds analyses of counterfactuals', you might reply. And indeed, that is a response perfectly consistent with the spirit of dispositional essentialism. Possible-worlds analyses are the product of trying to get modal truths by having a large number of non-modal truths. If you have enough worlds which are not themselves intrinsically modal, then you get modality free. But dispositional essentialists are prepared to bring modally thick properties into the actual world, so arguably they need no supplementation from others (Handfield 2001: 487).

Fair enough. Just because DE seems at odds with it it does not mean DE is untenable. In any case, I still think there is motivation for at least trying to show that DE may be in line with the Lewis-Stalnaker semantics. (I guess we have nothing to lose if there is a way to show that DE is in line with it). First, DE is not incompatible with, say, Lewis' and Stalnaker's formal systems for counterfactuals[5]. What

[4]Handfield tells us that this argument is due to John Bigelow (1999) who attributes its genesis to Jonathan Bennett.

[5]Consider, for example, Lewis' semantics for counterfactuals (call it LC). The following is given in Sider (2010: chapter 8): An LC-model, $\mathcal{M}$, is an ordered triple $\langle \mathcal{W}, \preceq, \mathcal{F} \rangle$, where

- $\mathcal{W}$ is a nonempty set
- $\mathcal{F}$ is a two-place function that assigns either True or False to each sentence letter relative to each $w \in \mathcal{W}$
- $\preceq$ is a three-place relation over $\mathcal{W}$.

The valuation function for $\mathcal{M}$ and $\preceq$ satisfy the conditions that:

- for any x, y if $\preceq_x x$, then $x = y$.
- for any $w \in \mathcal{W}$: $\preceq_x$ is strongly connected in $\mathcal{W}$
- for any $w \in \mathcal{W}$: $\preceq_x x$ is transitive

The truth condition for $\boxright$ is: $\text{LV}_{\mathcal{M}}(\phi \boxright \psi, w) =$ True iff either ϕ is true in no worlds, or: there is some world, x, such that $\text{LV}_{\mathcal{M}}(\phi, x) =$ True and for all y, if $y \preceq_x x$, then $\text{LV}_{\mathcal{M}}(\phi \rightarrow \psi, y) = True$.

dispositional essentialists cannot accept is Lewis' argument about how to evaluate counterfactuals on the assumption of determinism, so it seems incompatible with his ranking of comparative overall similarity as presented in the previous chapter, which is part of Lewis' account of counterfactuals. If so, as Handfield himself puts it, one's *modus ponens* is another's *modus tollens*. According to Handfield, Bigelow presented the argument as an objection to dispositional essentialism. Yet, if the dispositional account of laws is correct, then this is a *reductio* of Lewis' account of counterfactuals. This is why Handfield tells us that there is at least something at stake for both dispositional essentialists and those who follow Lewis' account.

A solution to this problem is interesting by its own rights. But it may reveal itself useful to this discussion as well. After presenting Handfield's solution, I will draw a parallel between Handfield's solution to the problem and a certain libertarian view.

5.4 Space-invading properties and Handfield's solution

The ranking of overall comparative similarity is supposed to capture the idea that the respects of similarity and dissimilarity (concerning the laws of nature and particular matters of fact) are traded off against one another. In providing such a ranking we want to say that, if a certain proposition P is false at the actual world, on the assumption of determinism, the most similar P-world will be the one that is exactly like the actual up to the time to which P obtains, and then diverges as a result of a divergence miracle. Clearly, this is possible according to a Humean view, for if the laws of nature are different so are the particular matters of fact. Handfield's solution tries to accommodate this idea without allowing laws to be broken, and so without allowing divergence miracles. He proposes something perhaps even more exotic: a law-abiding miracle.

The "nearness relation" $\preceq$ does not make any reference whatsoever to laws. So, I do not see a reason to suppose that $\preceq$ is incompatible with DE-laws.

> I now introduce the concept of *a space-invading property-instance*: a property-instance which occurs spontaneously, without any preceding cause. Let X be a property instantiated spontaneously at the same time as event d. X has the following causal powers: (a) to inhibit the causal process through which the properties of d normally lead to e-events (there are at least two ways familiar from recent literature on dispositions in which this inhibition might occur: the space-invader may be a *fink* for the disposition associated with d-events, it may act so as to make them lose the disposition to cause e-events; alternatively, the space invader must be an *antidote* to the process whereby d-events lead to e-events, i.e., it may interfere with the causal chain between d-events and e-events; and (b) to cause, in conjunction with the properties of d, e'-events. (Handfield 2001: 488)

Phew! Let me try to explain what this all means in complete detail and how this helps replying to the objection to DE. Start with finks.

5.4.1 An interlude: finks and antidotes

Finks

Finkish dispositions are better grasped in the context of the counterexamples to the conditional analysis of dispositions (CA). Charlie Martin (1994), for example, famously presented a counterexample to (CA) that hinges on finkish dispositions. I follow Bird in saying that in order to understand finkish dispositions we need to note two of their characteristics. First of all, dispositions have duration; the process whereby a disposition manifests will take time. Birds gives as examples the poison's disposition to kill which in order to manifest needs first to interact with the metabolism of the victim. Or consider someone who has the disposition to be angry: "the irascible man may swift to anger but not literally instantaneously" (Bird 2007: 25). Second, things may have a certain disposition at time t but not at a different later time t', that is, dispositions may be gained or lost. "Some food may become infected with the bacterium *Clostridium botulinum* and thereby become poisonous. It can lose that disposition by cooking or irradiation" (Bird 2007: 25).

To sum up:

(1) The process whereby a disposition manifests will typically take time.

(2) Many dispositions may be gained or lost.

Now suppose a certain object o has the disposition to M in response to stimulus S. Suppose o receives stimulus S at t_1 and that the manifestation occurs at t_3 such that t_1 is earlier than t_3, that is, $t_1 < t_3$. However, suppose that o loses its disposition just after it receives the stimulus and in a way that it loses the disposition before t_3, say, at t_2. How is that possible? As dispositions may be gained or lost, it could be that o loses the disposition at a time before t_3, such as t_2. Therefore, o will not manifest its disposition at t_3 since it was lost at a time earlier than that, namely, at t_2. Now we have the ingredients to put forward a counterexample to the simple conditional analysis. Just to recapitulate:

(SCA) Something x is disposed to manifest M in response to stimulus S if, and only if, if x were to undergo stimulus S, x would give response M.

"Simple indeed - but false", as Lewis puts it (1997). Well, (1) and (2) allow us to build a scenario in which (SCA) fails. Let us consider the original counterexample presented by Martin who, as Lewis says, decisively refuted (SCA). Martin focuses on the left-to-right implication of (SCA):

(SCA←) If, if something x were to undergo stimulus S, x would give response M, then x is disposed to manifest M to stimulus S.

Suppose that a certain wire is "live" if it is disposed to conduct electrical current when touched by a conductor. And let us assume

(A) The wire is live.

(SCA) will tell us that the wire is disposed to be live if and only if "if it were touched by a conductor, it would conduct electrical current". Now consider the following scenario:

> **Electro-fink**: The wire referred to in (A) is connected to a machine, an electro-fink, which can provide itself with reliable information as to exactly when a wire connected to it is touched by a conductor. When such contact occurs the electro-fink reacts (instantaneously, we are supposing) by making the wire live for the duration of the contact. In the absence of contact the wire is dead. For example, at t, the wire is untouched by any conductor, at t_2 a conductor touches it, at t_3 it is untouched again. The wire is dead at t_1, live at t_2, and dead again at t_3. In sum, the electro-fink ensures that the wire is live when and only when a conductor touches it (Martin 1994: 2).

The original counterexample Martin presented shows why the left-to-right implication of (SCA) is false. Suppose that at t_1 the wire is not touched by a conductor. Thus, since the electro-fink ensures that the wire is live when and only when a conductor touches it, the wire is not live at t_1. However, the counterfactual "if the wire were touched by a conductor at t_1, it would conduct a current" is true, since the electro-fink would ensure that the wire would conduct a current if touched by a conductor. Thus, (SCA←) is false.

If we consider the electro-fink in its reverse cycle, we have a counterexample to (SCA→). That is, suppose that the wire is dead when and only when it is touched by a conductor. Suppose that at t_1 the wire is *not* touched by a conductor. Thus, the wire is not dead, that is, it is live. However, the counterfactual "if the wire were touched by a conductor at t_1, it would conduct a current" is false, since the electro-fink would ensure that the wire is dead whenever it is touched by a conductor. Thus, (SCA→) is false.

In sum, some object's dispositions are finkish because the object may lose the disposition after the stimulus occurs but before the manifestation can occur. In Martin's scenario the electro-fink is connected to the stimulus in a way that whenever the object receives the stimulus it also loses the disposition. But things do not need to be put in that way. It could be that the object loses its disposition after the stimulus occurs in a completely accidental way.

Antidotes

Finks are not the only problem to the conditional analysis. Lewis' (1997) for example has tried to wipe out the counterexamples to the conditional analysis by precluding finks. Finks may occur because the object may lose the disposition after the stimulus occur but before the manifestation can occur. Lewis argued that in order to rule out finkish dispositions the causal basis of the disposition needs to remain for a sufficient time, so that the manifestation can occur. His reformed analysis may be presented in the following way:

(LCA) Something x is disposed at time t to give manifestation M to stimulus S iff, for some intrinsic property B that x has at t for some time t' after t, if x

> were to undergo stimulus S at time t and retain property B until time t', S and x's having B would jointly be an x-complete cause of x's giving response to M.

An x-complete cause of y includes all the intrinsic properties of x which causally contribute to y's occurrence. (As Bird points out, this stipulation is required to rule out certain other finkish counterexamples).

I am not entirely sure how Lewis' reformed analysis may rule out Martin's case. Perhaps the idea is that the electro-fink interferes in the process of the wire conducting electrical current when touched by a conductor; and by doing so the causal basis of the disposition to conduct electrical current when touched is lost.

Nevertheless, Bird has argued that Lewis' reformed analysis does not rule out *antidotes*:

> Many dispositions have what I call antidotes. An object x is disposed to display reponse r under stimulus s. At time t it receives stimulus s and so in the normal course of things, at some later time t', x gives response r. The time gap between t and t' is what allows, in finkish cases, for the loss of a disposition. An antidote to the above dispositions would be something which, when applied before t', hs the effect of breaking the causal chain leading to r, so that r does not in fact occur. Thus one can ingest a lethal dose of poison, yet not die if a suitable antidote is administered soon enough. (For instance, the antidote to arsenic poisoning is dimercaprol, which,incidentally, is also known as British Anti-Lewisite.) I suggest that the existence of antidotes provides counter-examples to Lewis' analysis (Bird 1998: 228).

Let "fatally poison" mean disposed to kill if ingested. LCA tells us that this is so if the poison has the causal basis remained for a sufficient time. I may ingest the poison and yet survive, for I may ingest the poison at the same time I ingest an antidote. Of course, the poison is left unchanged and all its intrisinc properties (whatever they are) are left unchanged, so that the causal basis remains unchanged. If that is so, then it is not a finkish disposition. The object in question still has the disposition. The poison has the disposition to kill when ingested, and this disposition is not lost; contrary to finkish cases, the disposition is not lost because of the time gap between stimulus and manifestation. That is, in the case of antidotes "the disposition and its causal basis remain throughout"[6] (Bird 1998: 228).

[6]Lewis, however, was unconvinced that antidotes are counterexample to (CA) or (RCA). Rather,

5.4.2 Space-invaders

Now that I have briefly explained what finks and antidotes are we may come back to Handfield's point. He talks about space-invaders in a context where he considers the following chain of global events in a world w (a global event is a state of an entire world at a time):

(World w): a caused b caused c caused d caused e caused f.

Let us also consider his concept of a *space-invading property instance* again:

Space-invasion: Let X be a property instantiated spontaneously (without any preceding cause) at the same time as event d. X has the following causal powers: (a) to inhibit the causal process through which the properties of d normally lead to e-events; and (b) to cause, in conjunction with the properties of d, e'-events[7].

Handfield's suggestion is that a space invader is not incompatible with the laws of w. That is, he tells us that when the global event d occurs, it brings about the "miraculous" result of e'.

> But this is a completely law-abiding miracle. This instance of X is a space-invader, which occurs spontaneously; but that is not incompatible with the laws which govern X. The laws make no mention of how instances of X must be brought into being. Moreover, when X is instantiated as a component of a global [d]-event, the law-governed effect is to bring about e'-events (Handfield 2001: 489).

As Handfield points out, the doctrine of dispositional essentialism is compatible with the concept of a space-invading property instance. All that DE tells us is that

---

he considers them as counterexamples to simplistic analyses of covert dispositional locutions (such as covert dispositional property names, like "fragility" and "combustibility", or predicates, such as "fragile", combustible") into overt dispositional locutions (such as "the disposition to break when stressed" or "is disposed to M when S").

[7] Handfield himself considers an ambiguity in this formulation. "When X occurs, is the global event the mereological sum of d and X, or does the instantiation of X change the nature of the world so that d does not occur at all?" (Handfield 2001: 488-489). For ease of exposition I will admit the compossibility of d and X.

"the powers of a property are essential to it. It does not, however, assert about the causal *means* by which a property can come to be instantiated" (Handfield 2001: 489). If this is so, then perhaps the dispositionalist may allow that all of the actual properties could have space-invaders instances.

Back to our discussion, the idea is that the dispositional essentialist can replace the notion of a divergence miracle with the notion of a space-invader[8]. For example, assuming that I did not raise my hand at the actual world but that I could have, the most similar world to the actual will not be the one at which a divergence miracle occurs. Rather, it will be the one at which a space-invader occurs; perhaps a spontaneous firing of a neuron in my brain causing me to raise my hand contrary to what was expected to actually happen. "In this way, space-invaders can do their job of bringing about counterfactual antecedents in just as subtle a fashion as Lewisian miracles." (Handfield 2004: 412).

There is a very important point, however. Handfield's conception of space-invaders - contrary to Lewisian divergence miracles - is incompatible with DE laws being deterministic and universal in extent. For consider world w. Let P_d stand for d occurs, and let P_e stand for e occurs, such that the global event e includes that I do not raise my hand. Let us suppose that the laws of w are DE-deterministic and universal, so that they cover P_e and that $P_d(L, P_e)$ is input to L for P_e. Now, let us consider the closest world to w where I raise my hand. Call it $w*$. We want $w*$ to be just like w up until the time at which I raise my hand. That is, we want $w*$ to agree on $P_d(L, P_e)$. But if $w*$ agrees on $P_d(L, P_e)$, it will also agree on P_e. Thus,

[8] If we consider Lewis' ranking of priorities as presented in chapter 4, the dispositionalist will have something like this:

(1) It is of first importance to avoid big widespread, diverse space-invaders.

(2) It is of second importance to maximise the spatio-temporal region throughout which perfect match of particular fact prevails.

(3) It is of third importance to avoid even small, localised, simple space-invaders.

(4) It is of little or no importance to secure approximate similarity of particular fact, even in matters that concern us greatly.

$w*$ will not be a world where I raise my hand.

But isn't this a problem for Handfield's solution? After all, the main point was to defend that some counterfactuals are true on the assumption that laws are deterministic and universal. But if Handfield is right, then DE-laws cannot be deterministic and universal. Here is what he says in response to that:

> Well, in the strong sense in which it is usually meant, I am suggesting that it is open to the dispositional essentialist simply to deny the supposition that determinism is possible. This does not mean that dispositional essentialists cannot endorse some weakened sense of determinism. That sense would be that a world might be such that no space invaders and no chancy properties are ever instantiated. For example, a Newton-world might exist, and it might have no space-invaders instantiated in it. This world would satisfy the most hard-core Laplacean variety of determinism, but this fact about the world would not be a law in and of itself. It would be a contingent *de facto* regularity that no determinism-disrupting properties were ever instantiated in the world. That does not amount to determinism *de jure*. (Handfield 2001: 490)

Now, if laws can be interfered with by a space-invader, we need to say something about the logical form of Handfield's *deterministic* laws. First, if any dispositional properties are capable of space-invasion, then the laws of nature will be *ceteris paribus* laws, or at least they will have a *ceteris paribus* clause concerning the absence of space-invasion. Second, Handfield tells us that the law will require some sort of global formulation, that is "for the simple reason that otherwise factors extrinsic to the antecedent always get in the way and prevent from coming about" (Handfield 2001: 492). Finally, "these sorts of laws will work only for non-chancy dispositional properties" (Handfield 2001: 492).

The way I see how the possibility of space-invaders is compatible with Handfield's deterministic laws can be presented as follows. The question we must ask is this: what follows from these deterministic laws? It seems to me that Handfield will be committed to something like the following, where SI is the proposition that there is space-invasion:

(D*): If L is deterministic* then for any P that L covers and any $P_{t0}(L, P)$, $\Box((P_{t0} \wedge \mathcal{L} \wedge {\sim}SI) \supset P)$

That is, (D*) is consistent with accepting that "SI" is possible, and hence is consistent with space-invasion. This is a sense in which the laws of nature are

limited in extent; Handfield's reformed deterministic laws can be cast as a call for domain restriction.

In the next section, I shall use use a similar strategy to introduce the view that free will requires the laws of nature to be limited in extent. It is usually assumed that the theory of agent causation is not consistent with the laws of nature being deterministic and universal. What I suggest is that, if free will and determinism are incompatible and there is free will, the laws of nature do not have to be indeterministic. We can accept that they are limited in extent: they do not apply to agent-caused actions. This is because an agent-caused event can be conceived as spontaneous in the same way that space-invaders are spontaneous.

5.5 Agent causation

Handfield explains the inhibition of the causal process through which the properties of d normally lead to e-events in terms of a property that is instantiated spontaneously. To that I add something else: agent-causation. It seems that, just like finks and antidotes, (a) and (b) can be satisfied by an agent exercising her agent-causal powers - if there is agent-causation.

But what is the theory of agent causation? Roderick Chisholm remarkably introduced it as follows:

> we have a prerogative which some would attribute only to God: each of us, when we act, is a prime mover unmoved. In doing what we do, we cause certain events to happen, and nothing - or no one - causes us to cause those events to happen. (Quoted in Vihvelin 2013: 57)

Here, I shall discuss O'Connor's account because he is very clear about the fact that the metaphysical assumptions of agent-causation are *anti-Humean.* My aim is to use it as a working hypothesis to give more flesh to the parallels between space-invaders and agent causation. O'Connor provides an account of reasons explanations *and* the metaphysics of agent-causation, and here I will only focus on the latter[9].

[9]See O'Connor's chapter 5 of *Persons and Causes* (O'Connor 2000). For the general conditions "to explain an action in terms of an antecedent desire" see O'Connor (2000: 86).

5.5.1 The metaphysics of agent causation

O'Connor's account of the metaphysics of agent causal-powers springs from his understanding of causation in general: agent causation is just another species of the same causal relation[10]. The first assumption of O'Connor's theory concerns the general concept of causation. He accepts a non-reductive understanding of causation where the primitive idea of the causal production, or "causal oomph" in a more technical jargon, is at its heart (O'Connor 2000: 67). He tells us that G. E. M. Anscombe advocates the non-reductive understanding of causation in her inaugural lecture "Causality and determination":

> [C]ausality consists in the derivativeness of an effect from its causes. This is the core, the common feature, of causality in its various kinds. Effects derive from, arise out of, come of, their causes. For example, everyone will grant that physical parenthood is a causal relation. Here the derivation is material, by fission. Now analysis in terms of necessity or universality does not tell us of this derivedness of the effect; rather it forgets about that. For the necessity will be that of laws of nature; through it we shall be able to derive knowledge of the effect from knowledge of the cause, or vice versa, but that does not show us the cause as source of the effect. Causation, then, is not to be identified with necessitation. (Anscombe 1981: 67)

By adopting a non-reductive account O'Connor automatically rejects those accounts along Humean lines. (For instance, causation cannot be explained in terms of a chain of causally dependent events). As O'Connor says, "the acceptance of the irredutibility thesis leaves open more than one direction on the way towards a 'thick' theory of event causation" (O'Connor 2000: 69). He rejects the Drestke-Tooley-Armstrong approach for reasons that needn't concern us here and adopts a causal powers account instead, that is developed by R. Harré and E. H. Madden in *Causal Powers: a theory of natural necessity* (1975).

O'Connor tells us that "there is another species of the causal genus, involving the characteristic activity of purposive free agents" (O'Connor 2000: 72). According

[10]Even though I am introducing the notion of causation to present O'Connor's view, I am assuming in this chapter that properties are essentially such as to confer certain causal powers or dispositions. According to Handfield it seems that dispositions and causes are somehow "deeply connected", and "a reasonable heuristic assumption, therefore, is that a correct understanding of dispositions would shed light upon the nature of causation, and vice versa". (Handfield 2009: 9)

to him:

> Parallel to event causes, the distinctive capacities of agent causes ('active powers') are grounded in a property or set of properties. So any agent having the relevant internal properties will *have it directly within his power to* cause any of a range of states of intention delimited by internal and external circumstances. However, these properties function differently in the associated causal process. Instead of being associated with 'functions from circumstances to effects', they (in conjunction with appropriate circumstances) make possible the agent's producing an effect. These choice-enabling properties ground a different type of causal power or capacity - one that in suitable circumstances is freely exercised by the agent himself. (O'Connor 2000: 72)

The ontological commitments of his account are these: universals and particulars are basic ontological categories.

Concerning particulars, and specifically agents, there are two general metaphysical assumptions. First, it is required that particulars continue to exist through time by being entirely located at each moment of their existence. That is, agent causation is committed to endurantism, as opposed to perdurantism. So it is possible that particulars can cause in a way that does not consist in causation by states or events. Also, it is not that all ordinary things that do not have temporal parts (and so are wholly present when they exist) may be considered "agents"; following Thomas Reid, O'Connor says that if something is an agent, then he is able to "represent possible courses of action to himself and have desires and beliefs concerning those alternatives (O'Connor 2000: 72). Second, the powers of agents must not be reducible to the powers of their micro-physical constituents, so that the powers of agents are "emergent". Concerning properties, "we require that they be universals that have essentially their dispositional tendencies" (O'Connor 2002: 73). Clearly, all the assumptions are very disputable, but they are not incompatible with dispositional essentialism.

There are two main differences between agent and event causation. The first and most obvious one concerns the causal *relata*. While the *relata* of event causation are events, the *relatum* of the "cause side" of agent- ausation is an enduring substance. Agent-causation is, as Ginet says, "the notion of a causal relation whose relatum on the cause side is not any event but the agent as such" (Ginet 1990: 12). There is an analogy here with our ordinary talk that may support such a distinction (Lowe

2001). We make statements concerning event-causation such as "the explosion of the bomb caused the collapse of the bridge", but we may also say something such as "the bomber caused the collapse of the bridge" (Lowe 2001: 2). In the latter example, the verb "to cause" takes as its grammatical subject a noun-phrase that might refer to an enduring substance, whereas in the former it might refer to a particular event. And since particular events and enduring substances belong to quite different ontological categories, it could be that the verb "to cause" has two different senses when referring to particular events and enduring substances. In any case, if O'Connor's view is right, the locution "s is the [agent]-cause of e" is to be interpreted literately, in the sense that the agent herself is the cause of e, and it cannot be analysed in terms of "Agent s caused event e if and only if there was some event, x, such that x involved s and x caused e".

The second difference concerns the idea of causal production or "causal oomph". With respect to event-causation, O'Connor says that it can be thought of as "functions from circumstances to effects". But agent-causal power cannot be thought in this way. It is not that someone agent-causes her coming to have a certain intention by being in the appropriate circumstances. Rather, the agent bears a set of essentially dispositional properties that can be thought of as choice-enabling, in the sense that if the agent has those properties, she is able to manifest the distinctive agent-causal power. The manifestation of such a power depends, of course, on the agent herself and not only on the circumstances where she is in.

5.5.2 Agent causation and space-invasion

Let us briefly go back to Handfield's world w, where the chain of global events from the dawn of time to the present goes like this:

a caused b caused c caused d caused e.

Let P_d stand for d occurs, and let P_e stand for e occurs, and suppose that the global event e does not include the event that I raise my arm. If laws are DE-deterministic and cover P_e, if another possible world w' agree on the laws and on

whether $P_d(L, P_e)$, they agree on whether P_e. Now consider what O'Connor tells us.

> One important feature of agent-causal power is that it is not directed to any particular effects. Instead, it confers upon an agent a power to cause a certain type of event within the agent: the coming to be of a state of intention to carry out some act, thereby resolving a state of uncertainty about which action to undertake. (For ease of exposition, I shall hereafter speak of "causing an intention," which is to be understood as shorthand for "causing an event which is the coming to be of a state of intention.") This power is multi-valent, capable of being exercised towards any of a plurality of options that are in view for the agent. We may call the causing of this intentional state a 'decision' and suppose that in the usual case it is a triggering event, initiating the chain of events constituting a wider observable action. (O'Connor 2009: 12)

The idea is that by manifesting the power distinctive to agent-causation, the agent causes an intention; we can say, for example, that the agent causes her coming to have a certain intention. Suppose that this occurs at the same time as event d, so that it initiates "the chain of events constituting a wider observable action" (O'Connor 2009: 12). Thus, the agent (a) inhibits the causal process through which the properties of d lead to e-events and (b) causes e'-events (the global e' event will include the agent raising her arm).

Handfield tells us that a space-invader is some sort of spontaneous event: "a property-instance which occurs spontaneously, without any preceding cause" (Handfield 2001: 488). In this sense, it is the event of the agent causing an intention. Consider an agent-caused event, such as

Agent-caused event (ACE): Pedro caused event e.

Chisholm raised the following question: "what, now, of that event - the event which is his thus causing e to happen? [...] Shall we say it that was not caused by anything?" (quoted in O'Connor 1995: 184). O'Connor's answer to this question is that an event such as (ACE) cannot be caused (O'Connor 2000: 58). If O'Connor is right, then events such as (ACE) are by their own nature uncaus*able*. They pop into existence and interfere in the natural order[11].

[11] To give a bit of historical context, Chisholm's worry was that if we answer "no" to the question of whether an agent causes the complex event (ACE), we cannot hold the agent responsible for his

But if O'Connor is right about the fact that (ACE) cannot be caused, the worry is this: how can the agent have control over his decision if the agent herself cannot cause the complex agent-caused event? His idea is that, once we admit that agent-caused events are by their own nature uncausable, the question is "ill framed, resulting from a failure to understand the distinctive nature of such an event. Agent-causal events are intrisically actions - the exercising of control over one's behavior. It is senseless to demand some further means of controlling this exercise of control" (O'Connor 2000: 59).

If the above is right, O'Connor provided a solution to a long-standing objection to agent causation. Now, this is certainly controversial, and it is not my aim to

---

causing e to happen. And he thought that if the agent is not responsible for (ACE), then the agent is not responsible for e either. Given this worry, Chisholm thought that the best answer to the first question is "yes", that is, that the event of the agent causing e to happen is caused by the agent. In other words, according to Chisholm, (ACE) is itself caused by the agent. Now, what about the event that (ACE) is caused by the agent? Of course, Chisholm's answer is that this event has to be caused by the agent as well, and so on, *ad infinitum*. To have a better grasp of this, consider Chisholm's account of agent-causation where he introduces a basic undefined locution: "he makes it happen that... in the endeavour to make it happen that...", where the gaps are to be filled in by sentences. Chisholm uses "**M**–..." in order to abbreviate the undefined locution and stipulates that it obeys the following axiom schemata:

(A1) Any instance of "**M**–..." implies $\exists p \exists q \mathbf{M}(p, q)$"

(A2) Any instance of "**M**–..." implies the corresponding instance of "—-".

(A3) Any instance of "**M**–..." implies its reiteration "**M**(**M**–...)...".

Chisholm introduced the axiom schema (A3) precisely to avoid the problem of explaining what causes (ACE). For O'Connor's objection to Chisholm, see O'Connor (2000: chapter 3, section 3.4).

O'Connor, of course, disagrees. As I said, he thinks that agent-caused events are by their own nature uncausable. The reason is this: consider a complex event such as

Complex event (CE): e causes d.

What does cause such a complex event? According to him, (CE) may be caused only indirectly. If c causes e, then c will cause the complex event (CE). However, if the *relatum* of the "cause side" is an enduring substance, say, Pedro, then nothing may cause it. For there is no clear meaning of something causing a substance *qua* substance.

provide a defence of agent causation. My main interest here is in the consequences of his view with respect to the dispositionalist conception of laws.

Remember that Handfield's reformed deterministic laws will cover situations where no space-invaders obtain. If agent causation can be thought of as space-invasion because agent-caused events are uncausable (at least according to O'Connor's view), what happens when agent-caused actions obtain? If we consider O'Connor's view, they will be constituted by essentialist dispositional properties (because agents bear a set of essentially dispositional properties). However, these properties are not like the properties that are "functions from circumstances to effects". In this sense, when an agent-caused action obtains, it will not be covered by the universal generalisation that supervenes on the dispositions that are "functions from circumstances to effects". Just like space-invaders, the reformed deterministic DE laws will not cover agent-caused events. What we have, thus, is something like this, where AC is the proposition that there is agent causation. Let P be the proposition that I do not raise my arm.

1. $\Box(AC \supset \Diamond(P_{t0} \wedge \mathcal{L} \wedge {\sim}P))$

2. $\Diamond AC$

3. $\Diamond \Diamond (P_{t0} \wedge \mathcal{L} \wedge {\sim}P)$ (from 1, 2, modal logic)

   Hence,

4. $\Diamond(P_{t0} \wedge \mathcal{L} \wedge {\sim}P)$ (from 3, S4).

The idea of introducing (1) is to cash out the idea that agent causation (just like space-invaders) may interfere in the natural order (Although I am not saying that agents can change, say, water into wine). It is clear that 2 and 1 are not consistent with DE laws being deterministic *and* universal (for the same reason that Sehon's interventionist God is inconsistent with the standard view of determinism). On the other hand, if we claim that the laws of nature are limited in extent because they do not apply to agent-caused events, we have

- $\Box((P_{t0} \wedge \mathcal{L} \wedge {\sim}AC) \supset P)$,

which is consistent with accepting that "AC" is possible, and hence is consistent with agent causation. Again, this is a sense in which the laws of nature are limited in extent, not that they are indeterministic.

Now, the formulation above may sound too implausible because it seems that we are requiring the laws to be silent about our actions. This is certainly a legitimate worry. But I do not think that I am requiring anything too implausible. What I am saying is just that, since dispositional essentialism is compatible with space-invasion, a formulation of the laws will - as Handfield himself writes - "require very strong restrictions on the possible interference of occurrences. Law statements will consequently refer only to highly idealised situations, which almost never obtain in actual practice" (Handfield 2001: 491). So, I think that it is not too surprising that these highly idealised situations will rule out interference by agents, especially in the case of agent causation. As we shall see in the interlude of the last chapter, this has some consequences with respect to the luck argument.

5.6 Final remarks

The dispositional account of laws justifies one of the premises of the consequence argument, namely, that the laws are not up to us. And the reason is that they are metaphysically necessary. I also have argued that, if dispositional essentialism is compatible with space-invasion, and if agent-caused events are uncausable, then DE laws will need to have a *ceteris paribus* clause with respect to the interference of agents. The view I just sketched has limitations, of course, because I did not argue for the view that events such as (ACE) cannot be caused. But I believe that this shows an alternative to the idea that, if determinism rules out free will, the laws of nature have to be indeterministic. I suspect that this requirement of laws being "indeterministic" is the one that generates worries about indeterminism and control, especially with respect to the luck argument. In any case, if we accept the dispositional account of laws and the add-ons we need for the truth of determinism, does it mean that there is no free will? That depends not only on the premise that the laws are not up us, but also on the premise about the past. And this is what I

will discuss in the next chapter.

Chapter 6

Can we change the past?

In order to reply to Warfield's charge of modal fallacy, proponents of the consequence argument need not only to show that its premises are true, but that they are necessarily true. This problem has been just partly responded since I have shown that $\Box_M \mathbf{N}\mathcal{L}$ and $\Box_M \mathbf{M}\mathcal{L}$ follow from the dispositionalist view of laws. The plausibility of incompatibilism depends thus on a successful reply to the main objection to the premise that no one has or ever had any choice about the past: Campbell's no past objection.

One intuitive picture in line with this premise being necessarily true is the garden of forking paths. One of the ways to spell this out is in terms of the idea that an agent's future is open, whereas the past and the present are not. That is, although an agent may have a choice about the future, she (1) cannot have a choice about the past (2) nor the present.

But how should (1) and (2) be formulated in a more precise way? Following the discussion on chapter 3, I will make use of the "no-choice" operator in order to draw up (1) and (2). I formulate the theses of the necessity of the past and the necessity of the present in terms of the "$\mathbf{M}$" operator. I fill a gap in the literature on this discussion: although it has been already argued that the necessity of the present and the necessity of past allow us to reply to the no past objection, no one (as far as I know) has formulated those theses in terms of an operator in which the conclusion of the consequence argument follows from its premises.

In section 6.1 I present the back and forth of the discussion on the necessity

of the past. In section 6.2 I formulate the theses of the necessity of the present ("NN" for short) and the necessity of the past ("NP" for short). After that I show that, contrary to Campbell's claim, they do not entail fatalism. Then I argue that $\mathbf{M}P_{t0}(L, P)$ is necessarily true given the assumption that there is no time earlier than t_0. This argument holds water if the necessity of the present is true. However, Bailey (2012) has provided counterexamples to it. In the final section I consider Finch's solution (2013a, 2013b) to this problem.

6.1 Campbell's no past objection

Following Warfield's considerations, Campbell has objected to the consequence argument by attacking premise 5. His first (well noticed) point is that the most obvious arguments presented in support of premise 5 are not cogent. There are two sorts of argument identified by Campbell. The first one goes like this:

(a) $\mathbf{N}P_{t0}$ because P_{t0} is a true proposition about the past and "no one can change the past" (Campbell 2007: 107).

The above argument is clearly invalid. It rests on the following invalid inference rule; where "PP" stands for "it has at some time been the case that P", the argument goes as follows.

1. PP

   Thus,

2. $\mathbf{N}P$

Let P be the proposition that I apply to study philosophy at Durham University. It is true that it has at some time been the case that P. And it is also true that, *right now*, I have no choice about whether P is true. However, just because I do not have a choice about whether P is true now it does not follow that I did not have a choice about whether P. I could have applied to study at a different university. Thus, 2 does not follow from 1.

Of course there is a straightforward response to that. For the premise the incompatibilist relies on is about a state of the world before any agents existed. The premise of the consequence argument is about the *remote* past. The argument for premise 5 can be better formulated, thus, in the following way:

(b) $\mathbf{N}P_{t0}(L, P)$ because $P_{t0}(L, P)$ is a proposition about the remote past.

Even so, Campbell argues that (b) will not do. The thesis that the laws of nature are deterministic does not need to involve a proposition about the *remote* past. We can think of a possible world at which the laws of nature are deterministic and agents existed throughout its history. Consider the following scenario:

> **[No past]** Consider, for instance, the possible world W. Suppose that W is a determined world such that some adult person exists at every instant. Thus, W has no remote past. At its first moment of existence lived Adam, an adult person with all the knowledge, powers, and abilities necessary for moral responsibility. Shortly after Adam comes Eve, and the rest is history. For each of the propositions that comprise W, someone is such that he has, or had, a choice about whether that proposition is true – at least there is no reason to doubt this claim. The Third Argument [that is, the modal consequence argument] is not a general argument for incompatibilism. At most, the Third Argument proves the weaker claim that persons cannot have free will in determined worlds with a remote past (Campbell 2007: 109).

If incompatibilism is to be formulated in terms of the strict conditional, then the no past scenario seems to show that the consequence argument fails to be an argument for incompatibilism. Put in that way, this looks like a boring (although correct) technical point. But Campbell does not put it in that way. He thinks the scenario tells against a weaker formulation of "incompatibilism" as well, that is, the thesis that determinism actually precludes the existence of free will:

> given only the Weaker Argument, we should judge that Adam is free in W but that Eve is not free. Yet it is hard to see how Adam and Eve differ in important respects. Good arguments for incompatibilism should expose the tension between the thesis of determinism and the free will thesis. If the Weaker Argument is the best that the incompatibilist has to offer, it remains a mystery why it cannot show that Adam lacks free will in world W. (Campbell 2007: 110)

Anthony Brueckner (2008) has tried to fix the problem by introducing a "$\mathbf{N}$" operator with times in it. Brueckner employs the operator "$\mathbf{N}_t P$", which stands for

"P and no one has any choice at t about whether P" (Brueckner 2008: 11). And he puts forward the same argument by mimicking the original inference rules:

(α*) For all t: from $\Box P$ derive $\mathbf{N}_t P$.

(β*) For all t: from $\mathbf{N}_t P$ and $\mathbf{N}_t(P \supset Q)$ derive $\mathbf{N}_t Q$.

Now, let P_{0*} be a proposition that expresses the total state of the world at some time t_{0*} prior to t (as Brueckner says, t_{0*} could be one second prior to t) (2008: 11). Here is Brueckner's new argument (BCA for short):

1	$\Box((\mathcal{L} \wedge P_0(L,P)*) \supset P)$	det. def.
2	$\Box(\mathcal{L} \supset (P_0(L,P)* \supset P))$	exportation, 1
3	$\mathbf{N}_t(\mathcal{L} \supset (P_0(L,P)* \supset P))$	$\alpha*$, 2
4	$\mathbf{N}_t\mathcal{L}$	premise
5	$\mathbf{N}_t P_0(L,P)*$	premise
6	$\mathbf{N}_t(P_0(L,P)* \supset P)$	$\beta*$, 3, 4
7	$\mathbf{N}_t P$	$\beta*$, 5, 6

Obviously, the most striking difference between BCA and the original argument is that the latter uses "**N**...,t". Brueckner's point, it seems, is that the trick of BCA is to show that no one can perform a free action at a time *relative to which there is a past*. The "appropriate" input to L for P, namely, P_{t0}* expresses the total state of the world at some time prior to t. This has an advantage over the original formulation; for example, P_{t0}* may be replaced by a proposition that is true 5 minutes before t, whereas the original one cannot.

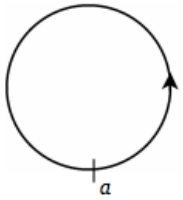
Even so, it is clear why BCA does not rule out the no past scenario. For suppose Adam performed action a at the first instant of w, say, t. For all BCA shows, Adam performed a freely at t.

Brueckner's idea here is to point out that, even though this is right, "no subsequent act of Adam's is free, and no subsequent act performed by any other human is free" (Brueckner 2008: 12). And that is perhaps the only limitation of BCA, one that seems quite unproblematic. After all, (historical) determinism spells out the idea that the past and the laws necessitate the future:

> On this conception of determinism, the assumption that there is no past relative to the time at which A is performed will quite naturally give rise to the consequence that there is no argument from the assumption of determinism to the conclusion that A is not free. (Brueckner 2008: 13)

Does it undermine Campbell's argument against weak incompatibilism as well? Campbell recognises that this might make some trouble for the argument against weak incompatibilism. And he has replied to Brueckner by presenting another scenario. His idea is to appeal to a situation at which time is circular; the instants are linearly ordered in a way that forms a circle, something like the conception of "eternal recurrence". See figure 1 (Mortari 2001).

Figure 6.1: Circular time



If a stands for *now*, the instants to the right constitute the future - but notice that if we go sufficiently far into the future, we will get to the past. Here is the scenario:

> **[Oscillating Adam]**. Suppose that there is a deterministic world, W*, where time is circular. In that world exists oscillating Adam. Oscillating Adam has always existed and will always continue to exist. He is in the grips of an everlasting, eternal recurrence. Oscillating Adam spends his time growing 'older' and getting 'younger'. He begins each cycle with powers comparable with the average 25 years old and eventually develops powers comparable with the average 50 years old. Then he slowly regresses back to the state at which he began, and the cycle starts all over again (2010: 72-3)

Campbell claims that this is a better scenario "because Adam no longer has an initial moment of existence. Still, Adam has no remote past; there is no necessity of the past to transfer onto Adam's future" (Campbell 2010: 73).

I will not discuss the question of whether Campbell's new scenario refutes Brueckner's argument. My aim is to use the definition of "**M**" presented in chapter 3 and show that the premise that no one has a choice about the input to L for P (boundary conditions, facts about the past, etc.) holds, so that even the new scenario should not worry incompatibilists.

6.2 Time-indexing in the "M" operator

The might-counterfactual sufficiency interpretation of "no-choice about" we considered in chapters 3 and 5 does not have times in it. And it is not clear at all how we can discuss Campbell's objection without time indexing. (After all you may think of P_{t0} as a proposition describing an event event, just like P). My strategy thus is to introduce the method of temporal arguments. And I shall introduce it for propositions as well. That is, instead of having the 0-ary predicate P we will use a monadic predicate $P()$ where the argument place is to be filled in by a term denoting a time. Let x, α and t range over, respectively, agents, action-types and times:

Definition 6.2.1. *$M_{t'}\phi_t$ if and only if $\phi_t \wedge \sim\exists x \exists \alpha \exists t'[Can(x, \alpha, t') \wedge (Does(x, \alpha, t') \diamond\!\!\rightarrow \sim\phi_t)]$*

Here is how Campbell's objection can be formulated now. If one wants to avoid the charge of modal fallacy, then one needs to show why $M_{t'}P_{t0}(L, P)$ is necessarily true. That is, necessarily, there is no agent x, no action α, no time t' such that x can do α at t' and if she were to do that at t', P_{t0} might be false, that is, P might be false at time t_0. But there seems to be a simple way to show that this is not necessarily true. Suppose that t_0 is the first instant of time at w in Campbell's original scenario. Let P_{t0} be a proposition about the total state of the world at t_0 in w and suppose that Adam did not raise his hand at t_0. According to Campbell, the possible world w is a world at which Adam can do something at t_0 (*e.g.,* raise his hand) such that, if he were to do it at t_0, P_{t0} might be false. Since t_0 includes Adam's not raising his hand, it is necessary that if he raises his hand, P_{t0} is false. Thus, if the antecedent is possible, it follows that if he were to raise his hand, P_{t0} might be false. That is, Campbell claims that it is possible that P_{t0} and $Can(adam, raise, t0) \wedge (Does(adam, raise, t0) \diamond\!\!\rightarrow \sim P_{t0})$.

If the purpose of the argument is to show that premise 5 is not necessarily true, the original scenario seems to do the trick. At least the way it was presented, if there is nothing wrong with it, then premise 5 is not necessarily true.

There is also an interesting aspect about the objection. If something happened

a while ago, I cannot do anything *now* to change it. This was the intuitive principle I formulated in the beginning of this chapter. I cannot do anything to change something that already happened. "**M**" helps us formulate this principle[1]:

(NP) $\Box\forall t\forall t'[t < t' \rightarrow (P_t \rightarrow {\sim}\exists x\exists\alpha[Can(x,\alpha,t') \wedge (Does(x,\alpha,t') \diamond\!\!\rightarrow {\sim}P_t))]$.

That is, necessarily, for any time t and any time t', if t is earlier than t', then if P is true at t, then no one can do anything at t' such that, if one were do it, P might be false at t.

Campbell's objection is interesting because it does not violate (NP) - the necessity of the past. This is so because Campbell is not claiming that Adam can raise his hand at some time t' later than t, the time at which P is true. Thus, (NP) alone will not help us in showing the truth of the premise that no one can change the past. As Roberto Loss has already argued (2009), if we want to establish that thesis, we need to appeal to another thesis: the necessity of the present.

Before I formulate this thesis, I would like to remind that it is in line with that intuitive "model": the garden of forking paths (whose name springs from Jorge Luis Borges' narrative). The idea is that the past and present are fixed; what happened and what happens cannot be changed. But the future is different; it is open and there many "alternative" futures, as it were.

What this model tries to capture is that it is only the future that is *up to us*. Neither the past nor the present are up to us. Of course, we can make choices in the present and they will affect the future. But that is different from saying that we can change how things are *now*.

It is this view, I believe, that motivates Loss' (2009) reply to Campbell. As far as I know, Loss was the first one to appeal to the necessity of the present in order to argue against Campbell.

> if I am running now, I cannot now do anything about the fact that I am running now. I can perhaps decide whether I will still be running in the next

[1]Notice that this has an advantage over the formulation of the so called "necessity of the past", at least as formulated by Arthur Prior: that is, if PP, then $\Box P$P. The problem is the one of whether the principle expresses the thesis of the necessity of the past in a mixed modal tense logic, and there are reasons to think it does not (White 1984: 60).

> hour, minute or second, but if it is true (if it is a fact, part of the actual world) that I am now running, this is something I cannot now make otherwise or prevent: *facta infecta fieri non possunt.* (Loss 2009: 67)

His formulation of the principle makes use of Brueckner's operator. But Brueckener did not provide an interpretation of "**N**" where we can prove its inference rules. I shall formulate the principle thus:

(NN) $\Box\forall t(P_t \rightarrow {\sim}\exists x\exists\alpha[Can(x,\alpha,t) \wedge (Does(x,\alpha,t) \diamond\!\!\rightarrow {\sim}P_t)]$

That is, necessarily, for every time t, if P is true at t, then no one can do anything at t such that, if one were to do it at t, P might be false at t.

We can use NN as a weapon to refute Campbell's objection. I will show that the assumption of the no past scenario leads to a contradiction.

1	$\Box\forall t(P_t \rightarrow {\sim}\exists x\exists\alpha[Can(x,\alpha,t) \wedge (Does(x,\alpha,t) \diamond\!\!\rightarrow {\sim}P_t)]$	NN
2	$\quad P_{t0} \wedge Can(adam,raise,t0) \wedge (Does(adam,raise,t0) \diamond\!\!\rightarrow {\sim}P_{t0})$	No past
3	$\quad P_{t0} \rightarrow {\sim}\exists x\exists\alpha[Can(x,a,t0) \wedge (Does(x,\alpha,t0) \diamond\!\!\rightarrow {\sim}P_{t0})]$	NN, 1
4	$\quad P_{t0}$	$\wedge$E, 2
5	$\quad {\sim}\exists x\exists\alpha[Can(x,a,t0) \wedge (Does(x,\alpha,t0) \diamond\!\!\rightarrow {\sim}P_{t0})]$	$\Rightarrow$E, 3, 4
6	$\quad \forall x\forall\alpha{\sim}[Can(x,a,t0) \wedge (Does(x,\alpha,t0) \diamond\!\!\rightarrow {\sim}P_{t0})]$	, 5
7	$\quad {\sim}[Can(adam,raise,t0) \wedge (Does(adam,raise,t0) \diamond\!\!\rightarrow {\sim}P_{t0})]$	$\forall$E, 6
8	$\quad Can(adam,raise,t0) \wedge (Does(adam,raise,t0) \diamond\!\!\rightarrow {\sim}P_{t0})$	$\wedge$E, 2
9	$\quad \bot$	
10	${\sim}[P_{t0} \wedge Can(adam,raise,t0) \wedge (Does(adam,raise,t0) \diamond\!\!\rightarrow {\sim}P_{t0})]$	$\neg$I, 2–9

I take this to be a refutation of Campbell's objection *if* the necessity of the present is a plausible thesis. Campbell has responded to Loss by claiming that the necessity of the present entails *fatalism*. Here, I will show that this formulation of (NN) does not entail fatalism. After refuting Campbell's point against Loss, I will go on to show how we can derive premise 5 of the consequence argument.

6.3 NP and NN do not entail fatalism

Campbell has argued that the necessity of the present and the necessity of the past entail fatalism. If this is the case, then it is a *reductio* to the response to the no past objection. Campbell's argument makes use of Loss' formulations of the principles, and Loss employs Bruckener's operator.

(γ) $\forall t \forall t' (t < t' \supset (P_t \supset \mathbf{N}_{t'} P_t)$

(δ) $\forall t (P_t \supset \mathbf{N}_t P_t)$

Campbell tries to establish that no one is able to do otherwise given these principles.

> I prove the generalization on the basis of a single, arbitrary case. I show that I am unable now to get my nose pierced, given that it was and is and always will be false that I get my nose pierced. The person (me), time (now) and proposition (that Joe gets his nose pierced) are arbitrary. Thus, no one is ever able to do otherwise and fatalism is true. (Campbell 2010: 74)

The strategy is as follows. If he is able now to get his nose pierced before, say, 2020, then it must be possible that his nose be pierced at some particular moment in time, say, P_t (Campbell 2010). And if that is the case, that moment will be either in the past, or in the present, or in the future. Suppose that that moment will be in the past. Let t be any past time and n be the present moment.

1	$\sim P_t \supset \mathbf{N}_n \sim P_t$	(γ)
2	$\sim P_t$	Assumption
3	$\mathbf{N}_n \sim P_t$	$\Rightarrow$E, 1, 2

Thus, the moment cannot be in the past. Now suppose that it will be in the present.

1	$\sim P_n \supset \mathbf{N}_n \sim P_n$	(δ)
2	$\sim P_n$	Assumption
3	$\mathbf{N}_n \sim P_n$	$\Rightarrow$E, 1, 2

Finally, if he is able now to get his nose pierced in the future, there must be some possible future moment of time when his nose is pierced. However, let $t*$ stand for any possible future moment.

1	$\sim P_{t*} \supset \mathbf{N}_{t*}\sim P_{t*}$	(δ)
2	$\sim P_{t*}$	Assumption
3	$\mathbf{N}_{t*}\sim P_{t*}$	$\Rightarrow$E, 1, 2

Campbell thinks that this is a *reductio* to the suggestion that an argument for incompatibilism should employ both γ and δ. Fortunately, he is not right.

Consider definition 6.2.1. Following Campbell I will characterise fatalism as the thesis that "no one can do otherwise":

(Fatalism) $\mathbf{M}_{t'}\phi_t$, that is $\phi_t \wedge \sim\exists x\exists\alpha\exists t'[Can(x,\alpha,t') \wedge (Does(x,\alpha,t') \diamond\!\!\rightarrow \sim\phi_t)]$.

This thesis is not established. The only thing he shows is that, if ϕ is true at a particular time t, then no one can do anything at that particular time t such that, if one were to do it, ϕ might be false at t. He *does not* show that no one can do anything at another time t' such that, if one were to do it at t', ϕ might be false at t. In other words, Campbell needs to show that *for every* t, no one has any choice about whether ϕ. The only thing he shows is that no one has a choice about whether the proposition is true *at the time* (or after the time) the proposition is true.

Thus, the principles do not entail fatalism, and Campbell's objection fails[2].

6.4 Arguing for premise 5

Now I want to argue that premise $\mathbf{N}P_{t0}$ is necessarily true. Notice that the sentence P_{t0} serves as a dummy which we can replace by any true sentence about the past or present we like (provided it is about the total state of the world).

(I) P_{t0}

By (NP), we have:

(II) $P_{t0} \rightarrow \sim\exists x\exists\alpha\exists_{t'>t_0}[Can(x,\alpha,t') \wedge (Does(x,\alpha,t') \diamond\!\!\rightarrow \sim P_{t0})]$

Given that, *by stipulation*, there is no time t' *earlier* than t_0, we can infer:

[2]After writing this section, I discovered that Loss (2010) replied to Campbell as well.

(III) $P_{t0} \rightarrow {\sim}\exists x \exists \alpha \exists_{t' < t_0} [Can(x, \alpha, t') \wedge (Does(x, \alpha, t') \diamond\!\!\rightarrow {\sim}P_{t0})]$

By (NN):

(IV) $P_{t0} \rightarrow {\sim}\exists x \exists \alpha [Can(x, \alpha, t_0) \wedge (Does(x, \alpha, t_0) \diamond\!\!\rightarrow {\sim}P_{t0})]$

Thus we can infer:

(V) $P_{t0} \rightarrow {\sim}\exists x \exists \alpha \exists t' [Can(x, \alpha, t') \wedge (Does(x, \alpha, t') \diamond\!\!\rightarrow {\sim}P_{t0})]$

Finally, from (I) and (V) we can deduce:

(VI) $P_{t_0} \wedge {\sim}\exists x \exists \alpha \exists t' [Can(x, \alpha, t') \wedge (Does(x, \alpha, t') \diamond\!\!\rightarrow {\sim}P_{t0})]$

Notice that (VI) is equivalent to $\mathbf{M}_{t'} P_{t0}$. Since w is any arbitrary world, we can derive the necessity of $\mathbf{M}_{t'} P_{t0}$, that is, $\Box \mathbf{M}_{t'} P_{t0}$. Thus, premise 5 follows.

It is high time that we considered the plausibility of (NN) and (NP). I will only make some very brief comments.

I am assuming without argument that there is no question of having a choice about a proposition about the past by means of time travel or the like. After all, Campbell's counterexamples did not show that agents may be free in a world with backwards causation, time travel, etc. On the contrary, his argument is compatible with the necessity of the past as formulated in this chapter. So, it seems that this premise should not worry us.

How about (NN)? This is certainly the most controversial one. I do think that this is quite plausible. If only the future is up to us, we can only have a choice about the future. This is formulated as follows: I can do something such that, if I were to do it at t, P might be false at $t' > t$. Now we need to do more to motivate this thesis. I will do this by considering Finch's defence of the principle (in the context of a reply to Bailey's objection to the necessity of the present).

6.5 Bailey's objection to the necessity of the present

Bailey has presented three counterexamples to the necessity of the present. Here I shall consider what I take to be the most interesting one:

> The Instantaneous Chooser. I exist for but a moment (call it t): no more, no less. And when I exist, I freely perform an action; I choose, say, to consider the question of whether I will dream. [I consider the question of whether I will dream] is a fact about t, but I had a choice in the matter. And surely I didn't have a choice in the matter before—or after—my one moment of existence. So I had a choice in the matter at t. As before, we could, if we liked, add that indeterminism is true in at least this sense: for every proposition x expressing the complete state of the world at any time up to but not including t, the conjunction of x with the laws of nature is compatible with [I do not consider the question of whether I will dream]. (Bailey 2012: 366)

Bailey claims that, for all he can tell, this is a possible case. In that scenario, he chooses to consider the question of whether he will dream. Let P_t be the proposition that he chooses to dream at t. Still, he can do something at t such that, if he were to do that at t, P might be false at t. If this is right, then (NN) is false and so the strategy to derive $\mathbf{M}P_{t0}$ fails.

Bailey anticipates one objection to that counterexample. The idea is that the instantaneous chooser requires instantaneous causation. "That is, it can't be that any cause (a mental episode, say) exists for but an instant and is simultaneous with its effect (an action, say)" (Bailey 2012: 368). His reply to that objection is that the impossibility of instantaneous causation is less than obvious[3]. "And if arguments for incompatibilism must rest on such abstract considerations about causation, this does not bode well for them" (Bailey 2012: 368).

There are two responses to Bailey's objection to the necessity of the present. The first - and more plausible one - is Finch's strategy, where she defines "abstract time" and shows that (NN) follows from it. The second one - considered by Bailey himself - has more costs and requires to appeal to the necessity of origin.

6.5.1 Abstract times and Finch's trans-temporality thesis

It seems to me that the best way out for incompatibilists is to follow Finch's strategy in showing what she calls "the trans-temporality thesis". However, Finch's argument depends on some technical machinery. She follows Alvin Plantinga (1976) in

[3]Bailey tells us that philosophers such as Gordam (2004), Huemer and Kovitz (2003) have at least endorsed the possibility of instantaneous causation.

constructing modal claims as claims about possible worlds, where possible worlds are *maximal possible states of affairs*. Here is a brief paraphrase of it (Plantinga 1976: Sect. II. 1; Finch 2013b section 3). A state of affairs is possible if it might obtain and actual if it obtains. A state of affairs O obtains if and only if some object instantiates some property or stands in some relation to itself or something else. A state of affairs O includes a state of affairs O' if it is not possible for O to obtain and O' to fail to obtain; and O precludes O' if it is not possible that both obtain. O is a maximal state of affairs if and only if for every state of affairs O', O either includes or precludes O'.

Now we turn to the definition of abstract time:

> Abstract times might fruitfully be thought of as present-tense maximal states of affairs. Intuitively, and very roughly, a present-tense maximal state of affairs is a total state of the world at an instant, minus all of the past- and future-tense truths. More rigorously: Say that a state of affairs O is future directed just in case either O's obtaining entails that some contingent thing will exist or O's obtaining entails that no contingent thing will exist; and then define a past-directed state of affairs in the obviously parallel way. Then a state of affairs O is present-tense maximal if and only if, for every atomic state of affairs O' that is neither future-directed nor past-directed, either O includes O' or O precludes O' (Finch and Rea 2008: 10).

Finally, this is the trans-temporality thesis to be shown (T'T):

(T'T): Necessarily, for any agent x, any act a, any time t, and any time t', if (i) x performs a at t and (ii) it is up to x at t' whether x performs a at t, then $\sim(t = t')$.

An important aspect of this formulation is that it has a double time index: one being the time at which the agent performs the action and the other one being the time at which the action is up to the agent[4]. In other words, there is a transition from how-x-is-at-t' to how-x-is-at-t: x at t' has the property of being up to her to perform a at t and the property of being up to her to refrain from performing a at

[4]Finch seems to follow Lehrer when he says that "statements affirming that a person can do something have a double time index, one time reference being to the time at which the person has the capability, and the second being to the time of action" (Lehrer 1976: 243). See also van Inwagen (1983: 231 footnote 12) for a nice comment on this.

t. But by performing a at t, x lacks the latter property. Finch gives a necessary condition for an action being up to an agent:

(A): Necessarily, if it is up to an agent x at some time t' whether x performs an act a at time t, then x at t' is both (i) able to perform a at t and (ii) able to refrain from performing a at t.

From (A), it follows that:

(MP): Necessarily, if it is up to an agent x at some time t' whether x performs an act a at time t, then (i) it is possible at t' that x performs a at t and (ii) it is possible at t' that x refrains from performing a at t.

> With this, it is worthwhile to remind ourselves that if x performs a at t, t includes x's performing a at t. Indeed, given the definition of a time, it is logically impossible that t not include x's performing a, and so it is logically impossible that t includes x's refraining from performing a at t. What this means, then, is that if it is possible, at t', that x refrains from performing a at t, it is possible, at t', that t not obtain. And, by extension, this means that it is logically possible, at t', that some time $t*$ obtains, where $t*$ is some time that precludes x's performing a. (Finch 2013b: 478-9)

With this argument Finch claims that

(TP): Necessarily, if (i) it is possible at t' that x performs a at t and (ii) it is possible at t' that x refrains from performing a at t, then $\sim(t = t')$

follows. "And, if we consider (A) alongside (MP) and (TP), it is clear that the trans-temporality thesis (T'T) follows" (Finch 2013b: 479).

Here is how I see her argument. The idea is to show that if x performs a at t, then x is not able at t to refrain from performing a at t. Suppose that x in fact performs a at t. Given the definition of a time, it is not possible for t to obtain and x not to a at t. This is just to say that it is not possible for t to obtain and for x to refrain from performing a at t. Thus, x at t is not able to refrain from performing a at t. However, x at t' is able to refrain from performing a at t. Hence, t and t' are not identical. (T'T) and (NN) are just two different ways to express the same idea (cf. Finch 2013a: footnote 34). If the trans-temporality thesis is plausible, so is the necessity of the present. For if an agent has a choice about some proposition, the

agent has the ability to render this proposition false at some time different from the one at which the proposition is true. Finch's point is that her formulation gives a more detailed explanation as to why the necessity of the present is true.

The purposes of presenting Finch's argument for the necessity of the present was to show that the thesis is motivated by a certain definition of abstract times, and so that the necessity of the present is not an unjustified principle. There is, however, a different strategy that incompatibilists may take, although one that has some costs.

6.5.2 The necessity of origin and essentialist incompatibilism

The point of this strategy is to defend a certain assumption about the past in order to solve Campbell's problem; namely, that every actual person x is such that, *necessarily*, if x freely performs an action, there is a time prior to x's first moment of existence. Bailey considers this strategy, and correctly warns us that it has some costs. In any case, I think it is worth considering this strategy for the following reason.

My defence of the other premise of the consequence argument depends on a certain *essentialist* assumption, an essentialism concerning properties, that fundamental, natural properties are *essentially dispositional*. And this allowed us to show the necessary truth of $\mathbf{M}\mathcal{L}$ and $\mathbf{N}\mathcal{L}$. The idea here is to appeal to essentialism concerning individuals, namely, the necessity of origin. The reason is that the thesis that "every person is such that they cannot freely perform an action if there is no time prior to their existence" is an essentialist one, and I tend to agree with Lowe's suggestion that essentialist theses cannot be deduced from premises which do not *already include* an essentialist thesis (Lowe 2002: 107). The combination of a thesis concerning essentialism about properties (dispositional essentialism) and a thesis concerning essentialism about individuals (the necessity of origin) justifies a very strong form of incompatibilism: perhaps this could be called "essentialist incompatibilism".

The task of such a view is to justify the following claim:

(Temporal origin): every actual person x is such that, necessarily, if x freely performs an action, there is a time t prior to x's first moment of existence.

The essentialist thesis that we can appeal to (and that Bailey himself has also considered) is the necessity of origin. According to the necessity of origin, if something originated from a certain source in the actual world, then it also originates from the same source in every possible world in which it exists. Saul Kripke (1980: 114-16) has claimed that every human person essentially has the parents she has. Suppose this is the case. Then someone who follows Kripke could argue as follows (Bailey 2013: 363).

O1 Every actual human person s is such that necessarily, if s exists and s has parents, s has the parents s in fact has. (Premise, Origin Essentialism)

O2 Therefore: every actual human person s is such that necessarily, if s exists, there are some xs such that the xs are s's parents. (From O1)

O3 Necessarily, for any xs and any y, if the xs are the parents of y, then there is a time prior to y's first moment of existence. (Premise)

O4 Therefore: every actual human person s is such that necessarily, if s exists, there is a time prior to y's first moment of existence. (From O2, O3)

Now, premises O1 and O3 are quite controversial. But since I am here just interested in the consequences of such a view, let us suppose they are true. Given the argument is valid, if the premises are true it is also sound. Even so, there is still a problem, for O4 still does not justify the premise that $\Box\mathbf{M}P_{t0}(L, P)$. The reason is that the modality in O4 is *de re*, while in $\Box\mathbf{M}P_{t0}(L, P)$ it is *de dicto*. (One possible solution, nevertheless, is to appeal to the Barcan formula: $\forall x \Box \phi \rightarrow \Box \forall x \phi$).

The problem with such a view is that it cannot justify incompatibilism given the assumption that the laws of nature are universal in extent and deterministic. It needs to appeal to a very controversial thesis that is independent of our concerns about determinism. For this reason I believe that the necessity of the present - and also Finch's trans-temporality thesis - is a better response to Campbell's objection.

6.6 Final remarks

The main goal of this chapter was to put forward an argument for the premise that no one can change the past. One contribution to the current discussion was to formulate the theses in terms of an operator whose interpretation allows us to deduce the main inference rules in the consequence argument. Furthermore, it allowed us to show why the theses of the necessity of the present and the necessity of the past do not entail fatalism. This makes a strong case for premise 5, provided one finds the necessity of the present a plausible, or that Finch's trans-temporality thesis is plausible. If this is right, and given the dispositionalist view of laws, incompatibilism is true, which leaves the question as to whether indeterminism is compatible with free will to be answered.

Chapter 7

The *Mind* (and luck) argument

In chapter 2 I argued that the standard formulation of the problem of free will depends on certain add-ons about laws. Even so, I have assumed these add-ons throughout the following chapters of the thesis in order to look at the the conceptual relationship between determinism and free will with respect to the traditional problem. A Humean conception of laws allows the compatibilist to deny one of the premises of the consequence argument. Things are different for the dispositional account of laws, however. If laws are DE-deterministic and universal in extent, and the arguments presented in chapters 5 and 6 are cogent, then incompatibilism is true.

Nevertheless, remember that the *Mind* argument - just like the consequence argument - can be formulated in terms of a (β)-like transfer rule. It seems that accepting the validity of a (β)-like transfer rule boils down to accepting the validity of the *Mind* argument. But if the *Mind* argument is cogent (and determinism is incompatible with free will), there cannot be any room for free will. Here, I will argue that accepting the validity of the consequence argument does not commit one to accepting the cogency of the *Mind* argument. I will argue that there are many problems with the standard formulation of the argument, and that there are reasons to think that (β) is not pivotal to the *Mind* argument. After arguing that the main worry about the compatibility of indeterminism[1] and free will is not about accepting

[1]For ease of exposition I am still assuming that "indeterminism" here means that "determinism

a (β)-like transfer rule, I conclude with some brief remarks about the explanatory formulation of the luck argument. The chapter is structured as follows.

First, I consider some attempts to fill in the details of the *Mind* argument, as since Graham's objections (2010) it is not clear whether the argument is well-formulated. I look over the back and forth of the discussion surrounding the *Mind* argument, and argue that it fails to establish the incompatibility of free will and indeterminism. I begin with van Inwagen's original formulation and then discuss Finch and Warfield's response (1998) to the argument. Dana Nelkin (2001), however, convincingly argued that Finch and Warfield's response fails. After this exposition of the debate concerning the cogency of the *Mind* argument, I argue that we do not have good reasons for accepting its cogency. The overall conclusion is that the *Mind* argument formulated in terms of a (β)-like rule fails. After that I turn to the luck argument and show that there is a gap in one of its premises.

7.1 Van Inwagen's formulation of the *Mind* Argument

In *An Essay on Free will* van Inwagen sums up the argument as follows: "a free act is an act one has a choice about; but no one has any choice about that which is undetermined". Remember that van Inwagen defends this claim by asking us to imagine a mechanism the salient features of which are a red light, a green light, and a button. If someone presses the button of the mechanism, one of the lights will flash, but it is *undetermined* which. "Now suppose that you must press the button on this mechanism. Have you any choice about which of the lights will flash? It seems obvious that you have no choice about this" (van Inwagen 1983: 142).

Now consider the following scenario:

> THIEF: Let us consider the case of a hardened thief who, as our story begins, is in the act of lifting the lid of the poor-box in a little country church. He

is not true". In this sense, if the laws of nature are indeterministic, then indeterminism is true. But if indeterminism is true, it follows that either laws are indeterministic or that they are limited in extent.

> sneers and curses when he sees what a pathetically small sum it contains. Still, business is business: he reaches for the money. Suddenly there flashes before his mind's eye a picture of the face of his dying mother and he remembers the promise he made to her by her deathbed always to be honest and upright. This is not the first occasion on which he has had such a vision while performing some mean act of theft, but he has always disregarded it. This time, however, he does not disregard it. Instead, he thinks the matter over carefully and decides not to take the money. Acting on this decision, he leaves the church empty-handed. (van Inwagen 1983: 127-28)

An important assumption about the argument is that van Inwagen adopts the standard theory of action in which, very roughly, something counts as an intentional action if it is non-deviantly caused by the mental states and events that rationalise the action (such as beliefs and desires). The standard theory also provides a causal account of reasons explanation. In THIEF we have an explanation telling us why the thief left the church empty-handed by revealing the thief's reason for performing the action. According to the standard theory, there is a causal connection between the *explanandum* - the thief leaving the church empty-handed - and the *explanans* - the desires and beliefs of the thief that do the explaining.

In the scenario, the thief's beliefs and desires caused the event of his refraining from robbing the poor-box. In the argument, let DB stand for the proposition that at t the thief has both a desire to keep the promise he made to his dying mother and a belief that the best way to do this would be to refrain from robbing the poor-box. Let R stand for the proposition that at t' the thief refrains from robbing the poor-box. Remember that the basic idea of the *Mind* argument is that no one can have a choice about something that is undetermined. Here, it is undetermined whether R follows from DB. This gives the following argument.

1. $\mathbf{N}(DB \supset R)$

2. $\mathbf{N}DB$

   Thus,

3. $\mathbf{N}R$ (from 1, 2, and β)

Suppose that the scenario about the red-green device justifies the claim that "no one has a choice about that which is undetermined". Since it is undetermined

whether R follows DB, premise 1 is true. Van Inwagen asks us to imagine that the green-red device is "hooked up" to the thief's brain

> in such a way that, if it flashes green he will steal the money and if it flashes red he will repent and depart; and we may suppose that his coming to be in a state of uncertainty about whether to steal or to repent has the effect of pressing the button. It should be clear that the thief has no choice about whether to steal or to repent. (van Inwagen 1983: 143)

If the device is "functionally equivalent" (van Inwagen 1983: 143) to the thief's brain, then it seems he has no choice about that which is undetermined. And van Inwagen's point here, it seems, is that what drives this intuition is not the device itself, but the indeterministic outcome involved in it. (I will come back to this point later).

With respect to 2, it is also assumed that the thief has no choice about whether DB is true. For example, Franklin says that this "assumption is harmless as the thief could only have had a choice about DB [...] if he performed some earlier action which itself would have been brought about by yet earlier mental states. We could then raise the same questions about this earlier action and these earlier mental states. Someone might again insist that the agent had a choice about these still earlier metal states. But this cannot go on forever and we will eventually discover the thief's 'initial' mental states for which he had no choice" (Franklin 2011: footnote 35).

7.1.1 A libertarian solution: Finch and Warfield

The problem with van Inwagen's formulation is that (β) is too controversial. Although I have argued that (β) is valid if one is willing to accept Stalnaker's theory, it is pretty much consensual that proponents of the debate accept McKay and Johnson's counterexample. In other words, (β) is invalid given definition 3.7.1 according to Lewis' theory[2]. Well, at least Finch and Warfield's solution to the argument depends on (β) being invalid. Their strategy hinges on McKay and Johnson's influential counterexample to (β) so that they focus their criticism on "improved"

[2]Just to refresh the reader's memory, the definition is this: $\mathbf{N}\phi$ iff $\phi \wedge \sim\exists x\exists\alpha[Can(x,\alpha) \wedge (Does(x,\alpha) \boxdot\!\!\rightarrow \sim\phi)]$.

formulations of the *Mind* argument that are parallel to the consequence argument. Their goal is to defend the cogency of the consequence argument while denying the cogency of the *Mind* argument.

Let us consider the formulation of the consequence argument based on (β-2), which is a valid inference rule on the Lewis-Stalnaker semantics for counterfactuals given definition 3.7.1.

($\beta - 2$) $\Box(\phi \supset \psi), \mathbf{N}\phi \vdash \mathbf{N}\psi$

1. $\Box((P_{t0}(L, P) \wedge \mathcal{L}) \supset P)$
2. $\mathbf{N}(P_{t0}(L, P) \wedge \mathcal{L})$

   Thus,

3. $\mathbf{N}P$ (from 1, 2 and β-2)

Nevertheless, if the same rule is used in order to formulate the *Mind* argument, Finch and Warfield argue, we will have a valid but (clearly) unsound argument:

1. $\Box(DB \supset R)$
2. $\mathbf{N}DB$

   Thus,

3. $\mathbf{N}R$ (from 1, 2 and β-2)

Premise 1 is false. In THIEF there is the assumption that the event DB describes causes (but does not necessitate) the event R describes. As a result, they argue that since (β)'s invalidity is (pretty much) uncontroversial and the original formulations are invalid, the best solution is to go on offering an improved version of the arguments. And although both arguments rest on a valid inference rule the improved *Mind* argument is undoubtedly unsound. As a result, libertarians have nothing to fear.

One might object that premise (2) in the improved consequence argument is stronger than the premises in the original formulation. After all, if Finch and

Warfield are assuming the successfulness of McKay and Johnson's counterexample, what the counterexample shows, for starters, is that agglomeration is invalid. As a result, $\mathbf{N}(P_{t0}(L, P) \wedge \mathcal{L})$ does not follow from $\mathbf{N}P_{t0}(L, P)$ and $\mathbf{N}\mathcal{L}$. However, Finch and Warfield mantain that this should not be a problem:

> While it is true that the premise of the improved Consequence argument is formally stronger than the pair of premises in van Inwagen's argument, we maintain that the core intuition motivating the acceptance of van Inwagen's premises likewise motivates the acceptance of our premise. This core intuition is, we maintain, the intuition that the past is fixed and beyond the power of human agents to affect in any way. P describes the state of the world at some time in the distant past (before any human agents existed). L is a conjunction of the laws of nature which, we presume, in addition to being inalterable by human agents, do not change over time. Thus the conjunction (P and L) offers a description of what might be called the "broad past"—the complete state of the world at a time in the distant past including the laws of nature. We maintain, in asserting our premise, that the broad past is fixed in just the way that van Inwagen maintains that the past is fixed (and that the laws are fixed). (Finch and Warfield 1998: 523)

Finch and Warfield also consider a formulation with the "$\mathbf{M}$" operator[3].

1. $\mathbf{M}(DB \supset R)$

2. $\mathbf{M}DB$

   Hence,

3. $\mathbf{M}R$ (from 1 and 2)

And they also argue that this will not do because premise 1 is false:

> The premise states that R follows from (is indeterministically caused by) DB and there is nothing anyone could do (or could have done) which might result (or might have resulted) in R's not following DB. But given that DB's causing R is a case of indeterministic causation, there clearly is something one could do that might result in R's not following DB. In particular, any action (including inaction) at all that one performed is such that it might have resulted in R's not following DB. (Finch and Warfield 1998: 526)

It seems to me, however, that the above needs a much more detailed justification. But I will not attempt to provide such a justification because I think Nelkin has refuted Finch and Warfield's solution.

[3] $\mathbf{M}\phi$ if and only if $\phi \wedge {\sim}\exists x \exists \alpha[Can(x, \alpha) \wedge (Does(x, \alpha) \diamond\!\!\rightarrow {\sim}\phi)]$.

7.1.2 Nelkin: why the libertarian solution fails

Finch and Warfield assumed that the *Mind* argument could not be sound if formulated in terms of (β-2). However, Nelkin (2001) interestingly showed that (β-2) allows one to put forward an improved version of the *Mind* argument as well:

1. $\Box((DB \wedge (DB \supset R)) \supset R)$
2. $\mathbf{N}(DB \wedge (DB \supset R))$

   Thus,

3. $\mathbf{N}R$ (1 and 2, $\beta - 2$)

The justification for the premise of this improved version is analogous with the justification for the premise of the improved consequence argument. Nelkin thinks that the sort of reason that supports the premise in Finch and Warfield's improved consequence argument is exactly the same as there is in support of (2):

> Just as we can think of the conjunction of P and L as the 'broad past,' we can think of the conjunction of DB and $(DB \supset R)$ as the 'broad nature of our mental states'. In other words, the conjunction of DB and $(DB \supset R)$ represents the nature of an agent's mental states, including (some of) their causal properties. And, just as in a deterministic world the broad past is not up to us, in an indeterministic world it is not up to us what the nature of our mental states is. (Nelkin 2001: 113)

Nelkin's point is that the improved *Mind* argument is as strong as the improved consequence argument. As she says, "if the core idea behind the consequence argument is right, then so is the core idea behind the *Mind* argument". It is possible that both arguments fail, but Nelkin claims that "as long as we understand the ultimate conclusion in the way van Inwagen does, it is difficult to see how they could" (Nelkin 2001: 114).

If the arguments are sound, then libertarians are in a bad spot, since both determinism and indeterminism are incompatible with free will[4]. However, I shall argue that this formulation of the *Mind* argument is far from showing the *incompatibility* of free will and indeterminism.

[4]Nelkin thinks that this leaves some room for the idea that libertarianists do not need to be "leeway theorists", that is, that acting freely does not require the ability to do otherwise. This is

7.1.3 Graham: the discussion is confused

According to Graham (2010), this discussion is confused. He has noticed that an evident problem in the discussion is the assumption that $\mathbf{N}(DB \supset R)$ correctly symbolises the claim that no one has, or ever had, any choice about whether R follows DB. As he reminds us, the fact that $\mathbf{N}(DB \supset R)$ is assumed throughout the discussion to symbolise the claim that R follows DB is made evident in the following passages:

> If an agent's act was caused but not determined by his prior inner state, and if nothing besides that inner state was causally relevant to that agent's act, then that agent had no choice about whether that inner state was *followed* by that act. (van Inwagen 1983: 149, our emphasis)
>
> Once DB occurs, given indeterminism, perhaps R will follow and perhaps it will not but since once DB occurs everything relevant to R's occurrence has taken place it seems clear that no one has a choice about R's *following* DB. That is, it appears to follow that N($DB \supset R$). (Finch and Warfield 1998: 518, our emphasis)
>
> Now, since R is an indeterministic consequence of DB, it seems that no one has a choice about whether or not R *follows* DB. That is, N($DB \supset R$). (Nelkin 2002: 109, our emphasis)

The problem is to translate into logical notation the following claim:

(&) R follows DB and no one has, or ever had, any choice about whether R follows DB.

But Graham tells us that $\mathbf{N}(DB \supset R)$ is not the right way of symbolising (&). This is so because the claim that R follows DB entails both DB and R - and consequently the material conditional $DB \supset R$. But $DB \supset R$ does not entail DB and R. Clearly, the claim that R follows DB is stronger than the material conditional $DB \supset R$ (cf. Graham 2010: 279-80).

What is then the right way to symbolise the claim that R follows DB? Graham claims that to say that R follows DB is to say that what R describes takes place a time later than what DB describes. Recall that DB is to be filled in by the

---

so because "it is strictly consistent with libertarianism that we act freely as long as our states of mind cause our actions and that indeterminism is true, even if we lack the ability to ensure that we do otherwise." (Nelkin 2001: 114).

proposition that at t the thief has the relevant belief/desire pair, and R by the proposition that at t' the thief refrains from robbing the poor-box. In other words, the event DB describes takes place at t while the event described by R takes place at t'. What we need is the additional claim that t' is later than t. Let F be the proposition that describes the event that t' is later than t. Then the right way to translate (&) is $\mathbf{N}((DB \wedge F) \wedge R)$.

Yet, if $\mathbf{N}((DB \wedge F) \wedge R)$ is the premise of the *Mind* argument, there is no need to employ rule (β) (or any similar β-like transfer rules) in the argument. After all we can put forward the argument like this:

(δ) : $\mathbf{N}(\phi \wedge \psi) \vdash \mathbf{N}\psi$

1. $\mathbf{N}((DB \wedge F) \wedge R)$

   Thus,

2. $\mathbf{N}R$ (from 1 and δ)

But if that is the argument, Graham says it is "blatantly question-begging" because "to be justified in accepting $N((DB \wedge F) \wedge R)$ one would have to be antecedently justified in accepting the conclusion of the argument, namely, **N**R" (Graham 2010: 280).

At this point one may wonder why we are spending so much time with this boring examination of the argument's formalisation. This passage is instructive:

> Why, you might ask, couldn't they all have just dispensed with talk of R's following DB and presented the *Mind* Argument explicitly in terms of the material conditional? Here is why. The *Mind* Argument's premises are offered as intuitively obvious. No arguments are offered in their support. The thief scenario is presented and it is simply asserted that it is obvious from the situation presented that the two premises are true. Given that this is how the premises are supported, it makes all the difference in the world what we take the ordinary language formulation of those premises to be because it is only in their ordinary language formulations that we can at all claim them to be intuitively obvious. (Graham 2010: 281)

If Graham is right, then the *Mind* argument fails to be a problem for libertarians who accept the validity of a (β)-like transfer rule. After all, the premises of the

argument formulated in terms of the material conditional are not, as Graham says, intuitively obvious.

I think that there some other difficulties with this formulation of the argument. Let us take a look at them.

Are the premises necessarily true?

The argument is supposed to be one for the *incompatibility* of indeterminism and free will. One of the ways to formulate it is as follows: "if determinism is not true, there is no free will". However, that conditional cannot be the material one. To see that, consider a hard determinist who thinks that incompatibilism is true and determinism is true. A hard determinist will accept the truth of the material conditional "if determinism is not true, then there is no free will", just because he thinks that the antecedent is false. If we want to establish the *incompatibility* of two theses, we cannot just show that the material conditional is true.

Now, what if the "if-then" of that conditional is the one of the strict implication? In that case, if indeterminism strictly implies that there is no free will, then free will and indeterminism are incompatible. However, in that case, the standard formulation of the *Mind* argument would be a modal fallacy. In order to get the conclusion that indeterminism strictly implies that there is no free will, we need the premises of the argument to be necessarily true. But no arguments are presented for the necessity of the premises. It seems to me that, given the assumption that indeterminism is true, at least one premise is not necessarily true, and which one is not depends on where indeterminism is "located" in this formulation of the argument.

If incompatibilism is true, then if there is free will, free will needs to be compatible with the laws of nature being indeterministic (assuming, of course, that they cover everything, but I shall ignore this complication for now). But *where* exactly indeterminism should be located? It cannot be that there is free will because some particles in a far, far away galaxy are not determined. Although libertarians disagree about where to posit the requisite of indeterminism, it must be somewhere in the aetiology of a free action. Now, in our discussion, we can understand DB as being undetermined ("undetermined" in the sense that laws are indeterministic

with respect to DB), or R being undetermined, or R being undetermined by DB. For our purposes, it does not really matter, for at least one premise will not be necessarily true.

Suppose, for example, that it is undetermined that R follows DB. If laws were deterministic and were to cover R, then DB and L would entail R. However, since they are indeterministic with respect to it, it is possible that L and DB and $\sim R$. If so, it follows that it is possible that $DB \wedge \sim R$. As a result, even if one takes $\mathbf{N}(DB \supset R)$ as intuitively obvious, the argument still does not prove the incompatibility of indeterminism and free will, for the very same reason that the original formulation of the consequence argument fails to be an argument for incompatibilism. If we need to defend that the premises are necessarily true, it needs to be the case that $\Box\mathbf{N}(DB \supset R)$. But $\Box\mathbf{N}(DB \supset R)$ entails $\Box(DB \supset R)$, which contradicts the assumption that it is possible that $DB \wedge \sim R$.

On the other hand, if we understand DB as being undetermined, then it possible that L and $\sim DB$, so that it is possible that $\sim DB$. If so, then $\Box\mathbf{N}DB$ is false, for $\Box\mathbf{N}DB$ entails $\Box DB$, which contradicts the assumption that it possible that $\sim DB$.

I am not saying that there is no other way to put forward an argument for the incompatibility of indeterminism and free will. I am just saying that this formulation fails to do so. Anyway, even if one gets rid of this problem, there is a more serious one concerning the notion of "undetermined". I attempt to show that this reveals that rule (β) is not really pivotal to the *Mind* argument.

What does "undetermined" mean?

Remember that van Inwagen defends the claim that "no one has any choice about what is undetermined" by appealing to the green-red device scenario. It is not totally clear to me, however, what "undetermined" means in this context. According to Franklin's formulation of van Inwagen's argument, if someone presses the device's button, "there is *an objective probability* of 1 that either the red light or green light will flash, but an *objective probability* of less than 1 that the red light will flash and *an objective probability* of less than 1 that the green light will flash" (Franklin 2011: 225, my emphases). So, according to Franklin's formulation, "undetermined"

here means that determinism is not true in the sense that the laws of nature are indeterministic with respect to the outcome of the device. In other words, the laws of nature do not always select a single outcome (for example, that the red light will flash), although they lay down at least a probability.

Now, suppose we distinguish the questions about extent and permissiveness (chapter 1). If determinism is not true, what follows is that, either the laws of nature are not deterministic, *or* that they are not universal. If we consider Franklin's formulation, "undetermined" means that the laws are not deterministic. But suppose that - as articulated in chapter 5 - the laws of nature do not cover free actions. If so, the scenario seems beside the point. What free will requires is that the actions of agents are outside the domain of the laws, not that they are indeterministic in Franklin's sense. The upshot is that even if free will is incompatible with laws being indeterministic for agents, free will is still possible as long as laws are limited in extent and do not cover actional-events.

Of course, there is a clear objection to the above point. For suppose that by "undetermined" we mean that the laws are limited in extent and do not cover the device. Even so, it seems we do not have a choice about this. We do not have a choice about whether the red or green light will flash, after all this is undetermined.

I remain unconvinced. I agree with Franklin that the example van Inwagen gives establishes far less than what he thinks. What he seems to be assuming is that, by pressing the button, the agent does all he can, and then just has to wait to see what the outcome is (Franklin 2011: 205), that is, the outcome comes about "indeterministically" or "randomly". In that case, of course, the agent has no control over the outcome. However, I do not see how this is supposed to follow from the claim that the laws do not cover such an outcome because they do not cover agent-caused events. If by exercising her agent-causal power the agent brings about R, then the agent exercises her control, even though the laws do not cover it.

In fact, Seth Shabo (2013) has made the same point when he claimed that rule (β) is not really central to the *Mind* argument:

> According to these libertarians, rational agents (*qua* substances) sometimes make a further causal contribution to their free actions, one that we overlook if we attend only to their beliefs, desires, and other mental states. Suppose for

> discussion's sake that the thief makes this further, "agent-causal" contribution to R. On this basis, these libertarians will say, it *is* up to the thief whether R occurs, notwithstanding that DB are the only *states* relevant to R's occurrence. If this appeal to agent-causation succeeds in explaining how R can be up to the thief, the libertarian can reject N(DB occurred ⊃ R occurred). (Shabo 2013: 295)

On the other hand, if agent causation is not the correct account of control, then agent-causalists will not be able to explain why $\mathbf{N}(DB \supset R)$ is false. In any case, the crucial question here is whether agent causation can correctly explain control.

Of course, "[t]he same point applies, *mutatis mutandis*, to libertarian opponents of the *Mind* Argument who reject agent-causation" (Shabo 2013: 295). Here is what Franklin, for example, says:

> [van Inwagen's] claim that the thief does not have a choice about whether R follows DB does not seem well-formed. To ask whether the thief had a choice about whether DB brought about R is, on the face of it, a rather awkward question since, according to the event-causal theory of action, DB's bringing about R just is the thief's exercising control over R. (Franklin 2011: 227)

So, the crucial question here is not whether accepting (β) commits one to accepting the cogency of the *Mind* argument. The crucial question is the one of whether we can successfully explain what it means to exercise control over an action[5].

We have three difficulties with the *Mind* argument. First, we do not have good reasons to think that the claim that "no one has a choice about whether R follows DB" can be symbolised in terms of the material conditional. And if we present the *Mind* argument explicitly in terms of the material conditional, it is not intuitively obvious that the premise $\mathbf{N}(DB \supset R)$ is true. Second, even if it is obvious that the material conditional is true, the standard formulation of the *Mind* argument is an

[5]Not surprisingly, there is disagreement among agent-causalists (such as Griffith 2010; O'Connor 2011) and event-causalists (such as Franklin 2011, 2014). Both theories are supposed to fulfil the requirement of control without assuming the truth of determinism. The disagreement however is about whether control can be "solely analyzed in terms of, or reduced to, states and events involving the agent" (Franklin 2014: 141). Event-causalists also include Wiggins (1973), Ekstrom (1993, 2000), Kane (1996, 1999). One recent argument against the event-causal approach that favours the agent-causal one is the disappearing agent objection (Pereboom 2007: 102). See also Franklin (2014), Runyan (2016), and Clarke (forthcoming).

ignoratio elenchi, for it does not establish the *incompatibility* of indeterminism and free will. In order to do that, we need to show that the premises are necessarily true, but the participants of the debate agree that it is possible that $DB \wedge \sim R$, so that the premise cannot be necessarily true. (And even if we say, on the other hand, that DB is undetermined, the first premise will not be necessarily true). Third, there is the problem of explaining what "undetermined" means in the context of this discussion. And what is crucial here is whether we can have a correct account of control even on the assumption that determinism is false. But it is far from being intuitively obvious that, just because laws are silent (or indeterministic) with respect to our actions, it follows that we have no choice about them.

The main worry that the *Mind* argument seems to pose - at least the way I understand it for the purposes of this thesis - is that we cannot simultaneously accept a (β)-like transfer rule while denying the cogency of the *Mind* argument. But if what I have considered here is correct, this not a problem at all. First, because it is implausible to think that the argument should be formulated in that particular way in terms of a (β)-like transfer rule. And second because even if there is no problem in formulating the argument in that way, it is far from obvious that its premises are true, let alone necessarily true. Because the main interest here is in the *compatibility* of indeterminism and free will, I conclude that *this* formulation of the *Mind* argument does not compel us to accept that free will *entails* determinism. This does not mean that there are no better formulations of the argument. Graham himself considers several different formulations of the argument only to conclude that they all fail, but none of them relies on a (β)-like transfer rule as discussed here, and none of them rely on premises that are necessarily true.

The main worry then is about whether the falsity of determinism rules out control, independently of whether the argument is formulated in terms of a (β)-like transfer rule. A correct answer to this worry, however, depends on a discussion of how an agent may have a choice "about that which is undetermined". And here there is a huge disagreement between libertarians, and I am not getting into the discussion of who is right (the agent-causalist or the event-causalist).

Anyway, the luck argument seems to raise this worry in a better way than the

Mind argument, even without relying on a transfer rule like (β). As I said, I am not getting into the debate between event and agent-causalists. But I think that the distinction in chapter 1 between extent and permissiveness shows a gap in Mele's formulation of the luck argument.

7.2 Interlude: The luck argument

The *Mind* argument and the luck argument may be taken to be the same argument, but here I have distinguished them because the formulations I am discussing are different. The worry about the *Mind* argument was a worry related to (β)-like transfer principles. Here I will focus instead on the explanatory formulation (Mele 2006). Let us discuss the argument in this schematic form (Franklin 2011):

1. If agent s performs action a at t in w and there is some world $w*$ that has the same laws and past, but in which s performs not-a, then there is nothing that accounts for the difference between world w in which s performed a and $w*$ in which s performed not-a.

2. If nothing accounts for this difference, then it is partly a matter of luck that s performed a in w and partly a matter of luck that s performed not-a in $w*$.

3. If an action is partly a matter of luck, then the action is not free

   Thus,

4. If s performs action a at t in w and there is some world $w*$ that shares the same laws and past, but in which s performs not-a, then both a and not-a are not free.

The worry about luck can be raised if we consider the thief scenario in the *Mind* argument. In that scenario, the thief has reasons for robbing the box (he wants some money, he is confident no one will catch him, etc.) and reasons for refraining from robbing the box (he made a promise to his mom, he wants to fulfill that promise, etc.). What happens in the story is that the thief decides not to rob the box at t. According to the standard libertarian, this implies that there is another world that

shares an initial segment up until t, the time of the decision, with the same laws of nature. That is, there is another world with the same laws and past where the thief decides to rob the box at t. However, since both worlds have the same laws and past, the thief has exactly the same mental states, reasons, deliberations, etc., prior to t. How is the decision up to him if there is *nothing* to explain this difference?

One of the ways to reply to that argument is by saying that "Mele is right that nothing can explain the occurrence or non-occurrence of a other than [the thief's] decision. But this is no argument against the libertarian, since it is exactly her position that we have this undetermined capacity to shape the future according to our own decisions" (van Miltenburg & Ometto forthcoming: 4).

Mele however has already replied to that objection:

> I have heard it said that what I am presenting as a problem for typical libertarians cannot possibly be a problem for them because their view entails that cross-world differences of the sort at issue are required for directly free action. But, of course, sometimes a philosopher's view entails something impossible [...] And the answer that it has to be possible because its possibility is required by typical libertarian views is a remarkably poor answer. (Mele 2014: 548)

I agree that this is a poor answer. Yet, I do not think that the libertarian needs to require this sort of cross-world difference for free will, at least not if we consider the view I articulated before. In chapter 5, I put forward an argument for the claim that agent causation (or at least O'Connor's account of agent causation) requires the laws of nature to be limited in extent, rather than being indeterministic. I argued that agent-causation is compatible with the laws of nature being deterministic if the laws have a "no agents intervening" clause. And I think this plays a role in the discussion concerning the conditional "necessarily, if determinism is not true, there is no free will", especially with respect to the first premise of Mele's formulation of the luck argument.

If the luck argument is an argument for the claim that free will *entails* determinism (that is, the contrapositive of "necessarily, if determinism is not true, then there is no free will"), there is a gap in it. The crucial point can be raised by the following question: does the antecedent of premise 1 follow from the claim that determinism is not true?

Remember that "determinism" here means that the laws of nature are both

universal in extent (that is, they cover everything) and are deterministic. However, if we say that laws are limited in extent in the sense I mentioned above, it does not follow that we need to accept that "agent s performs action a at t in w, and there is some other world $w*$ with the same laws and past in which s does otherwise". So, agent-causation (conceived as space-invasion), contrary to what Mele says, does not entail the impossible.

It seems to me that Mele's formulation presupposes that the laws of nature are indeterministic with respect to the event that agent s performs a. Clearly there are further implicit conditions in the antecedent if (1) is to follow from the assumption that the laws of nature are indeterministic. I take it that it is presupposed that a is the kind of happening that is governed by laws and that those laws that govern it are indeterministic with respect to it. In other words, let us say that P is an action of a person (agent s performs a). Laws L are indeterministic in the sense that for any P that L covers and any input $P_{t0}(L, P)$ and any possible worlds w and $w*$ in which $\mathcal{L}$, w and $w*$ agree on $P_{t0}(L, P)$ but not on whether P obtains. But, of course, this already presupposes that L covers P.

However, I argued that since dispositional essentialism is compatible with space-invasion, a formulation of the laws requires - as Handfield himself writes - "very strong restrictions on the possible interference of occurrences. Laws statements will consequently refer only to highly idealised situations, which almost never obtains in actual practice" (Handfield 2001: 491). And I argued that these laws would have to rule out the interference by agents as well, for agent-caused events (at least according to O'Connor's view) are uncausable and interfere in the natural order. If so, it cannot be that a is the sort of event governed by the laws.

The difference in this response is that I am not attacking premise 2:

2. If nothing accounts for the cross-world difference, then it is partly a matter of luck that s performed a in w and partly a matter of luck that s performed not-a in w.

Agent-causal libertarians, for example, think that what gives control over the action is not a certain sort of explanation but rather the agent-causal power exercised

by the agent. The reply then is to point out that the cross-world difference does not entail that the agent did not have control over a, so that a is not a matter of luck.

I do not claim that control requires lack of contrastive explanation. I am just saying that the agent-causal libertarian does not need to commit herself to the claim that the denial of determinism entails the antecedent of 1 in Mele's argument. But then again, the plausibility of such a reply will depend on agent causation (and, in particular, O'Connor's account of agent causation) being the right account of free will, which is something I did not consider in this thesis. While this is not a decisive refutation of the argument, at least shows that the libertarian does not need to be committed to the view that free will "requires the impossible".

7.3 Final remarks

I have argued the *Mind* argument - at least the standard formulation - is quite problematic. I also considered an implication of the view outlined in chapter 5 for the formulation of the luck argument. For if the "limited laws" account is right, the antecedent of 1 in the luck argument will not follow from the assumption that determinism is not true. If such an account of agent-causation is correct, the explanatory formulation of the luck argument will fail to establish the incompatibility of indeterminism and free will.

Chapter 8

Conclusion

My aim in this chapter is to give a very brief summary of the preceding discussion, focusing more on the big picture.

In chapters 1 and 2, I have formulated the problem of free will and determinism in terms of a dilemma against the view that there is free will. I have also distinguished several important questions with respect to "contingency", contending that the traditional problem of free will and determinism depends on the claim that the laws of nature are universal in extent and deterministic. However, as I argued in chapter 2, the claim that the laws are universal and deterministic is a controversial one that is just an add-on to the mainstream philosophical accounts of laws. I discussed, in particular, the "Humean" regularity theory, laws as relations among universals and the dispositional account. I also considered how the term "deterministic laws" should be defined relative to the philosophical accounts discussed there. My claim was that "Humeans" and powers advocates could adopt different definitions of "deterministic laws". I motivated the definitions by discussing Sehon's objection to the claim that deterministic laws are reliable.

In sum, if the arguments in chapter 2 are sound, there is nothing in the very nature of a law that implies that things could not happen other than the way they do consistent with the laws staying the same, nor even that probabilities need be fixed. Laws may be universal in extent and yet totally impermissive, and one may - or may not – have good independent arguments for these add-ons; but in all senses of "laws" surveyed that is just what these are: add-ons. It is important to keep this

questions separate from one another because they have consequences for the free will debate. In particular, I think that the distinction between extent and permissiveness plays an important role in the typical arguments for the incompatibility of free will and indeterminism. If determinism is not true, it does not follow that that the laws of nature are indeterministic.

Now, if determinism is true, then laws are deterministic *and* universal in extent. The compatibility problem is that of understanding whether determinism and free will are compatible. My focus was, of course, on the consequence argument. And because the bulk of the discussion concerned the modal formulation, I had to consider the debate over the (in)famous rule (β). Contrary to initial expectations, rule (β) and agglomeration are actually valid according to Stalnaker's theory (on the counterfactual sufficiency of "**N**"), although they are invalid on Lewis'. However, (β-M) is valid on Lewis' theory if we use an operator defined in terms of the "might-counterfactual". So, the discussion in chapter 3 showed that the main question is not that of whether the consequence argument is valid, but whether it is sound.

So, the question of whether the consequence argument is sound depends on what the laws of nature are. Different philosophical views have different approaches with respect to the compatibility of free will and determinism. If we assume the Humean view of laws - and Lewis' argument for miracles - then the premise that no one has a choice about the laws of nature is false. On the other hand, if we accept the dispositionalist conception of laws, this premise is necessarily true. If the arguments presented in chapter 5 are sound, then incompatibilists should take dispositional essentialism seriously.

Even if the premise that the laws of nature are not up us is necessarily true according to the dispositionalist conception of laws, the consequence argument relies on another premise: namely, that the past is not up to us. The real challenge here was to defend that this premise is necessarily true. Anyway, I believe that the defence of this premise is not really crucial for the purposes of the consequence argument because, even if Campbell is right, we can appeal to a restricted incompatibility thesis: namely, that free will and determinism are incompatible for agents relevantly like us (who do not live in a world where time is circular, or where there is no first

instant, etc.).

If the crucial question concerning the consequence argument is that about whether or not the laws of nature are up to us (and not the past), then it seems to me that the compatibility problem (the problem of whether determinism and free will are compatible) depends on what is the right account of laws. But does it depend on the view that laws govern? I think that this is an interesting question, and this is something that I would want to work on more. My impression is that Demarest's (2017) view that combines a best system account of lawhood with the thesis that the fundamental properties are essentially dispositional may also justify the the claim that the laws of nature are not up us. Also, a best system account does not require that the laws of nature cover everything, and it seems to me that this view could be articulated in a way not to cover actional events. This view would still be compatible with agent causation, and so would be a sort of non-governing libertarianism.

In chapter 7, I tried to show how important the distinction between extent and permissiveness is. My suspicion is that the problem with respect to indeterminism and free will arises because we presuppose that the laws govern or cover our actions, and are indeterministic with respect to it. This is even more evident in Mele's formulation of the luck argument. Clearly, that argument only goes through if we already presuppose that the laws govern everything and are indeterministic with respect to our actions. But what are the reasons for accepting that the laws govern everything else? Perhaps they do not need to. If so, premise 1 in the luck argument need not be true.

This brings us here. If my overall argument is sound, there is nothing in the very nature of a law that implies that there are no genuine alternative possibilities. Agents are able to do otherwise if we accept a Humean view of laws, where laws do not govern, or if we accept a view where deterministic laws obtain only on the assumption that there is no interference by agents. The only difference is that the Humean view is compatible with alternative possibilities even if determinism is true, while the dispositional essentialist view requires determinism not to be true (which, by the way, does not require laws to be indeterministic). However, showing that laws are consistent with various paths available to the agent does not show that

the agent has control, that is, that "the agent (and not some outside force or mere chance) selects which path will be the path into the future" (Fischer 1999: 99). The problem is that finding out what control is requires a trek farther down than the one I have traversed.

Appendix A

Better best system accounts

In chapter 2 I focused only on Lewis' BSA. But there are recent accounts of the BSA that make a departure from Lewis' formulation. The standard view of determinism has problems in accommodating recent accounts of laws in which laws of nature and laws of science go hand in hand. My goal in this interlude is to take a look at this issue in a bit more detail.

One of the attractive aspects of the BSA is that it does not appear to be an account of laws disconnected from science. Loewer, for example, is "very much attracted to the BSA because of the way it incorporates the criteria physicists use for counting generalizations and equations as expressing laws" (Loewer 2007: 313). Van Fraassen, however, has raised a problem concerning the connection between laws of science and laws of nature according to Lewis' BSA. The main controversy is on whether the BSA should depend on the language of natural properties. This, of course, needs clarification.

In formulating Lewis' BSA I pointed out that laws appear in the best systems that are deductively closed and whose non-logical vocabulary contains only predicates that express *perfectly natural* properties. Clearly, according to this formulation, Lewis' BSA depends on the notion of a perfectly natural property. But what does it mean?

Bird (2007) considers a distinction between two uses of the term "property". A philosopher may use the term "property" in a sense that "there is a property for every predicate or open sentence"; for example, she might say something like "the

property of being either blue or made of plastic". By contrast, there is a more restricted use of the term "property":

> For example, a scientist may discover or synthesize a hitherto unknown molecule. It would be natural to say that her next task is to "discover its properties". In that sense, its properties do not include "the property of being first synthesized on a Wednesday" or "the property of being ϕ" (where something is ϕ iff it is a member of the set (molecule, the Eiffel tower, the power set of the natural numbers) (Bird 2007: 9).

Lewis (1986) distinguishes the liberal use from the restricted one by distinguishing "non-natural" (or abundant) properties from "natural" (or sparse) properties. This is what he says about the distinction:

> Many philosophers are skeptical about the distinction between natural and gruesome properties. They think it illegitimate, unless it can somehow be drawn in terms that do not presuppose it. It is impossible to do that, I think, because we presuppose it constantly. Shall we say that natural properties are the ones that figure in laws of nature? - Not if we are going to use naturalness of properties when we draw the line between laws of nature and accidental regularities. Shall we say that they are the ones that figure in the content of thought? - Not if we are going to say that avoidance of gratuitous gruesomeness is part of what constitutes the correctness of an ascription of content. Shall we say that they are the ones whose instances are united by resemblance? - Not if we are going to say that resemblance is the sharing of natural properties. Unless we are prepared to forgo some of the uses of the distinction between natural and unnatural properties, we shall have no easy way to define it without circularity. That is no reason to reject the distinction. Rather, that is a reason to accept it - as primitive, if need be. (Lewis 1983: 344)

So, on Lewis' view not all properties are on a par; some of them - the perfectly natural properties - belong to a select group and are sparse. The distinction plays a crucial role in his formulation of the BSA because of the following problem.

Given any arbitrary system of true sentences S, we can always find a way to axiomatise S in a very simple way. Let F be a primitive predicate that applies to all and only things at worlds where S holds (Lewis, 1983: 42, cf. Loewer 1996: 185) and let S be axiomatised by $\forall x Fx$. Here is the problem. Suppose S is a system that includes all the truths of the actual world, so that it contains all true generalisations (including of course accidental generalisations). If we axiomatise S by the axiom $\forall x Fx$, we will have a very simple and informative system. But, of course, every true generalisation will be a law. As a result, the BSA will be unable

to distinguish laws from accidental true generalisations. This is how the distinction between natural and non-natural properties comes into play: "the simple predicates of the language in which systems are formulated (and in which their simplicity is evaluated) must express [perfectly] *natural* properties" (Loewer 1996: 185). That axiomatization of S will not be the best system because the predicate "F" does not express a perfectly natural property.

The distinction between natural and non-natural properties thus plays an important role in Lewis' formulation of the BSA. Without it, it seems the BSA will be unable to distinguish laws from accidental regularities. However, van Fraassen argues that there is a problem with Lewis' BSA precisely because it requires the simple predicates of the language in which systems are formulated to express such properties.

A.1 Van Fraassen's problem

Suppose that "at a certain point in history, all the primitive scientific predicates are natural ones" (van Fraassen 1989: 53). Van Fraassen invites us to imagine a scientist who enters the field with a new system, much simpler and stronger than the previous systems. However, this new system is formulated in a language whose basic predicates do not express perfectly natural properties. As van Fraassen points out, this cannot be a bad day for science. One should expect scientists to consider the new system the best one. And if that is the case, it could turn out that what scientists regard as laws of nature are not laws according to Lewis' view.

Loewer formulates the problem as follows:

> Suppose that FT is what Steven Weinberg calls "a final theory." FT maximally satisfies all the requirements that the tradition and practice of fundamental physics puts on a fundamental theory of the world. FT is true, simple, highly informative, comprehensive; FT reconciles relativity and quantum theory, explains statistical mechanical probabilities, and explains special science regularities, and so on. It does all this better than any alternative theory whether the alternative has ever been or ever will be thought up by anyone. There is no true theory that better than FT balances all these virtues. Even so it may turn out that some contingent generalizations/equations entailed by FT are not L-laws [laws according to Lewis]. Further, there may be L-laws that FT fails to entail (2007, 322)

The reason why this might happen is that FT - our "final" theory - may be formulated in a language - say $\mathcal{L}_F$ - whose basic predicates do not express perfectly natural properties. Now suppose that FT is translated into a language $\mathcal{L}_N$ whose basic predicates correspond to perfectly natural properties. As Loewer points out, it could be that, when translated into $\mathcal{L}_N$, FT is beaten out in the contest for best system by another system LT: when both systems need to be formulated in a language whose predicates express perfectly natural properties, such as $\mathcal{L}_N$, LT is stronger and simpler than the final theory FT. But it is also possible that, when both systems need to be formulated in $\mathcal{L}_F$, the original language of the final theory FT, FT is stronger and simpler than LT. So, which one is the best system after all? According to Lewis, LT is the best system, not the final theory. Loewer argues that this raises two problems, an epistemological problem and a metaphysical one.

The epistemological problem is that "knowing all the non-nomological contingent truths in every possible language isn't sufficient for knowing which truths are the laws. One would also have to know which predicates refer to Lewisian natural properties" (2007, 322). Another way to put the epistemological problem is in terms of epistemic accessibility (see Wheeler 2016), that is, Lewis' formulation seems to fail Earman's empiricist loyalty test (Earman 1986: 85). Earman formulates the problem in the following way:

Epistemic accessibility: for any two worlds w_1 and w_2, if w_1 and w_2 agree on all occurrent facts, then they agree on the laws of nature.

The only problem is to cash out the notion of occurrence, since it means different things to different philosophers who call themselves "Humean". How should one fill in this notion? There are many ways, "ranging from very strict empiricist (e.g., occurrent is actually observed) through less strict empiricist (e.g., occurrent is observable in principle), through much less empiricist (e.g., occurrent allows unobservable entities like quarks reached through expansive methods such as Glymourian boot- strapping or "inference to the best explanation")" (Cohen and Callender 2009: 9)[1]. Thus, it seems what distinguishes w_1 from w_2 is not something accessible via

[1]Armstrong's account of laws is usually thought to fail the test. Armstrong tells us that laws

standard scientific inquiry. On the other hand, the metaphysical problem is the one of "justifying why the aims of science should be coupled to the aims of Lewisian metaphysics" (Loewer 2007: 322).

In sum, it seems van Fraassen's problem undermines the attractive aspect that the BSA is connected to science. But since the problem is related to the dependence on the notion of perfectly natural properties, this is a good motivation to develop a version of the BSA without appealing to natural properties. That is exactly what Loewer, Callender and Cohen attempted to do.

A.2 Recent attempts to solve van Fraassen's problem

If one is willing to reject natural properties in order to formulate the BSA, then one option is to let scientists to decide what the acceptable terms are. This is Loewer's solution to the problem. According to him, science should tell us what is the right language. If scientists manage to carry out the "final theory", then one will be able to tell what the predicates of the best system express.

> Here is a way of thinking of the BSA in which Lewisian natural properties play no role. Consider the world w of all pairs $< L, T >$ of possible languages L and candidates for best systems of w T such that
>
> - T is formulated in L
> - T is true of w
> - T is a final theory for w (i.e., T is true and best satisfies the criteria of simplicity, informativeness, comprehensives, and whatever other conditions the scientific tradition places on a final theory for w.

---

of nature are necessary relations among first-order universals. Suppose the generalization "All F's are G's" is true at both worlds w_1 and w_2. F-ness and G-ness are taken to be first-order universals. According to Armstrong, if "All F's are G's" is a law of nature, then there is a second-order contingent relation holding between these two universals, call it "N(F,G)", such that N(F,G) entails the corresponding generalization. Since N is a contingent relation, it could turn out that N(F,G) entails "All F's are G's" at w_1 but not at w_2. Now, if "All F's are G's" is a law at w_1 but not at w_2, then w_1 and w_2 are distinct worlds. And what distinguishes w_1 from w_2 is the necessitation relation that only holds at w1. However, as Earman says, N(F,G) "is contingent on another category of facts which transcend the occurrent" (Earman 1986: 104).

> From all such pairs $< L, T >$ select the one (ones) $< FL, FT >$ that includes the best theory [where FL refers to the language of the final theory]. $< FL, FT >$ is the 'best of the best' and determines the laws of w. (Loewer 2007: 324)

Here is how to assess the best theory according to this view. First, there is a comparison of systems formulated with a particular language in order to determine the best system for that language. Second, from all such possible languages and best systems $< \mathcal{L}_1, T_1 >$, $< \mathcal{L}_2, T_2 >$, ... $< \mathcal{L}_n, T_n >$, there is a selection of the "best of the best" with respect to the conditions the scientific tradition places on a final theory. Loewer calls this view "the package deal account" (PDA) of laws because it identifies the laws and the nomological properties together (Loewer 2007: 324). This is why van Fraassen's problem should not be a problem. Furthermore, (PDA) can deal with Lewis' trivial system $\forall x Fx$, for Loewer may just claim that it lacks explanatory appeal from the point of view of the scientific community.

If we accept Cartwright's view (1989, 2009, 2010) that what we think of as the usual laws of physics may well be limited in extent, then the laws of nature according to Loewer's view account will be limited in extent.

A.3 Cohen and Callender's account

Just like PDA, Cohen and Callender's view allows the scientific community to decide the acceptable terms in which systems are formulated as well. There are three guiding ideas that motivate the development of their better best system account of laws (BBSA for short). That is, BBSA is supposed to satisfy the following three desiderata: the problem of immanent comparisons, the problem of epistemic accessibility and how to account for special science generalisations. I will focus for now on the first and the second and leave the third one for last.

With respect to epistemic accessibility I have already mentioned Earman's empiricist loyalty test. BBSA and PDA will agree on the fact that it is up to the scientific community to decide the acceptable terms in which systems are to be formulated. And I have already mentioned the concern about immanent comparison by considering the trivialisation worry raised by Lewis. But since Cohen and Callender think that the problem of immanent comparisons is prior to that worry

(2009: 6), a few comments are necessary. Following Quine's terminology (1970) they say that "simplicity is an immanent (defined relative to a system of basic kinds or basic predicates) rather than transcendent (defined independently of the system of basic kinds or basic predicates) notion" (Cohen and Callender 2009: 5)[2]. Cohen and Callender argue that simplicity, strength and balance are immanent rather than transcendent (Cohen and Callender 2009: 6). Consider simplicity, for instance. A language that contains "grue" and "bleen" as simple predicates (Goodman 1954) rather than "green" will count "All emeralds are green" as less simple than one that contains "green" as a simple predicate (cf. Loewer 1996: 109).

Cohen and Callender first consider (and then reject) what they call "stipulationism": when formulating the systems, scientists are allowed to a once-and-for-all stipulation of acceptable terms, so that all future systems must be formulated in terms of this "stipulated base". However, according to them,

> a once-and-for-all stipulationism offends against the anti-apriorism that inspires this theory of laws. As we just witnessed, any particular choice of X will preclude (what should be) live empirical possibilities. When we choose an X as our stipulated kind, we thereby remove X from the normal back-and-forth of scientific bartering (Cohen and Callender 2009: 18).

They offer an alternative view called "flexible stipulationism". The idea is that science should not be exempt from revision and evaluation.

> With respect to the problem of arbitrariness, the proponent of stipulative [BSA] can treat her stipulated fixed background (of kinds, observable, etc.) as a pro tanto, a posteriori, and defeasible assumption that is not insulated from empirical inquiry. The thought would be that, while some or other stipulated background is needed to carry out comparisons needed to fix [BSA's] laws, the background can itself be subject to rational revision on other occasions. (Cohenn and Callender: 2009: 20)

So what they propose instead is a relativised BSA, where there is no "transcendently best system (not fixed by nature, not stipulated once and for all by us)" (Cohen and Callender 2009: 21). The best systems are axiomatised relative to a specific choice of basic kinds K (to a specific choice of basic predicates P_K). For example, suppose there is a set $\mathcal{E}$ of economic predicates and a set $\mathcal{B}$ of biological

[2]This has been widely noticed by many authors such as (Lewis 1983: 366–368; 1986: 123–124; Earman 1986; van Fraassen 1989: 41–43, 51–55; Taylor 1993: 82; Loewer 1996: 109).

predicates. The general claims that represent laws of economics appear in the best system formulated in terms of economic predicates. The general claims that represent laws of biology appear in the best system formulated in terms of biological predicates.

The worry about immanent comparison is solved since systems can be compared with each other *provided they are formulated in the same language*. So, for example, systems formulated in terms of biological predicates can be compared with each other with respect to simplicity, strength and balance. The worry about epistemic accessibility is also settled as long as "the kinds in terms of which it is formulated are epistemically accessible" (Cohen and Callender 2009: 21).

As BBSA is an account that relativises laws to kinds it is natural to expect something to be said about what natural kinds are. Cohen and Callender interestingly suggest to adopt *explosive realism*:

> The answer in question is a proposed middle way between naive relativism and the idea that nature possesses a uniquely true carving up into kinds; on this view, the world permits possibly infinitely many distinct carvings up into kinds, each equally good from the perspective of nature itself, but differentially congenial and significant to us given the kinds of creatures we are, perceptual apparatus we have, and (potentially variable) matters we care about. Thus some sets of objects, although perfectly objective and well-defined, are not interesting to us, e.g., the set of the Eiffel tower, the two authors, and elm trees, whereas others are, e.g., the set of living creatures (Cohen and Callender 2009: 22).

And they emphasise that combining explosive realism with BBSA allows us to account for regularities in the special sciences. Notice that since on their view there is no transcendently best system, systems axiomatised relative to different choices of basic kinds (and so formulated in terms of different sets of predicates) cannot be compared with each other. So - for instance - a system axiomatised relative to a specific choice of basic kinds B (say, biological kinds) to a specific choice of biological predicates cannot be compared with a system axiomatised relative to basic kinds E formulated in terms of economic predicates. The pay-off is that this allows to count special science generalisations as law statements, even if their predicates do not pick out perfectly natural, fundamental properties[3].

[3]This, however, is controversial. See Backmann & Reutlinger (2014). For a defence of the

Lewis' BSA has problems in counting special science generalisations as law statements for two reasons. First, there is the requirement that systems need to be formulated in terms of predicates that express fundamental and perfectly natural properties, which is something that special science do not do. For example, imagine we construct a system in botany; it is unlikely that the basic predicates of such a system will pick out fundamental properties. Second, special science generalisations are usually thought of as allowing exceptions, and so "the fact that they are generally incomplete descriptions of reality imply that candidate special science laws will lose in terms of strength to candidate fundamental laws" (Cohen and Callender 2009: 23).

As one might expect, on such a view agents like us may be an exception to the laws that cover everything else in the world. After all the laws of nature depend on systems that are always constructed relative to a specific choice of basic kinds K (to a specific choice of basic predicates P_K). And it could be that these predicates fail to pick out properties concerning agents like us. So BBSA-laws may very well be limited in extent. On the other hand, it could be that some (physical, biological, etc.) laws are deterministic whereas the laws that cover agents are not. That is, laws may be permissive for agents while being impermissive for everything else.

There are many interesting aspects to consider, and the connection between the BBSA, explosive realism and free will is rather unexplored. It seems to me however that the BBSA will not have trouble with the claim that there is free will. My goal was just to point out the importance of the distinction between extent and permissiveness according to this view.

traditional BSA see Wheeler (2015).

Appendix B

Proofs of rules (α-M) and (β-M)

The following proof can be given on Lewis' system, where the following holds:

(L) $\phi \diamond\!\!\rightarrow \psi \dashv\vdash \sim(\phi \,\Box\!\!\rightarrow \sim\psi)$

B.1 α-M

We start proving (α-M):

1	$\Box\phi$	
2	ϕ	$E\Box$, 1
3	$\quad Can(s,a,t_1) \wedge (Does(s,a,t_1) \diamond\!\!\rightarrow \sim\phi)$	
4	$\quad Does(s,a,t_1) \diamond\!\!\rightarrow \sim\phi$	$\wedge$E, 3
5	$\quad\quad Does(s,a,t_1)$	
6	$\quad\quad \phi$	$E\Box$, 1
7	$\quad Does(s,a,t_1) \Box\!\!\rightarrow \phi$	$\Rightarrow$I, 5–6
8	$\quad Does(s,a,t_1) \diamond\!\!\rightarrow \sim\phi$	$\wedge$E, 3
9	$\quad \sim(Does(s,a,t_1) \Box\!\!\rightarrow \phi)$	L, 8
10	$\quad \bot$	
11	$\sim(Can(s,a,t_1) \wedge (Does(s,a,t_1) \diamond\!\!\rightarrow \sim L))$	$\neg$I, 3–11
12	$\forall x \sim(Can(x,a,t_1) \wedge (Does(x,a,t_1) \diamond\!\!\rightarrow \sim\phi))$	$\forall$I, 11
13	$\forall x \forall \alpha \sim(Can(x,\alpha,t_1) \wedge (Does(x,\alpha,t_1) \diamond\!\!\rightarrow \sim\phi))$	$\forall$I, 12
14	$\forall x \forall \alpha \forall t \sim(Can(x,\alpha,t) \wedge (Does(x,\alpha,t) \diamond\!\!\rightarrow \sim\phi))$	$\forall$I, 13
15	$\sim\exists x \exists \alpha \exists t[Can(x,\alpha,t) \wedge (Does(x,\alpha,t) \diamond\!\!\rightarrow \sim\phi)]$	DM, 10
16	$\phi \wedge \sim\exists x \exists \alpha \exists t[Can(x,\alpha,t) \wedge (Does(x,\alpha,t) \diamond\!\!\rightarrow \sim\phi)$	$\wedge$I, 2, 11

B.2 β-M

Now we prove (β-M):

1	$\mathbf{N}\phi$	
2	$\mathbf{N}(\phi \rightarrow \psi)$	
3	$\phi \wedge \sim\exists x\exists\alpha\exists t'[Can(x,\alpha,t') \wedge (Does(x,\alpha,t') \diamond\!\!\rightarrow \sim\phi)]$	def, 1
4	$\phi \rightarrow \psi \wedge \sim\exists x\exists\alpha\exists t'[Can(x,\alpha,t') \wedge (Does(x,\alpha,t') \diamond\!\!\rightarrow \sim(\phi \rightarrow \psi))]$	def, 1
5	ψ	$\Rightarrow$E, 1, 2
6	$\quad \sim(\psi \wedge \sim\exists x\exists\alpha\exists t'[Can(x,\alpha,t') \wedge (Does(x,\alpha,t') \diamond\!\!\rightarrow \sim\psi)]$	
7	$\quad \sim\psi \vee \sim\sim\exists x\exists\alpha\exists t'[Can(x,\alpha,t') \wedge (Does(x,\alpha,t') \diamond\!\!\rightarrow \sim\psi)]$	DM, 6
8	$\quad \sim\sim\exists x\exists\alpha\exists t'[Can(x,\alpha,t') \wedge (Does(x,\alpha,t') \diamond\!\!\rightarrow \sim\psi)$	DS, 5, 7
9	$\quad \exists x\exists\alpha\exists t'[Can(x,\alpha,t') \wedge (Does(x,\alpha,t') \diamond\!\!\rightarrow \sim\psi)$	$\neg$E, 8
10	$\quad Can(s,a,t1) \wedge (Does(s,a,t1) \diamond\!\!\rightarrow \sim\psi)$	$\exists$E, 9
11	$\quad \sim\exists x\exists\alpha\exists t'[Can(x,\alpha,t') \wedge Does(x,\alpha,t') \diamond\!\!\rightarrow \sim(\phi \rightarrow \psi))]$	$\wedge$E, 4
12	$\quad \sim[Can(s,a,t1) \wedge (Does(s,a,t1) \diamond\!\!\rightarrow \sim(\phi \rightarrow \psi))]$	$\forall$E, 11
13	$\quad \sim Can(s,a,t1) \vee \sim(Does(s,a,t1) \diamond\!\!\rightarrow \sim(\phi \rightarrow \psi))$	DM
14	$\quad Can(s,a,t1)$	$\wedge$E, 10
15	$\quad \sim(Does(s,a,t1) \diamond\!\!\rightarrow \sim(\phi \rightarrow \psi))$	DS, 13, 14
16	$\quad Does(s,a,t1) \boxdot\!\!\rightarrow (\phi \rightarrow \psi)$	L, 15
17	$\quad \sim[Can(s,a,t1) \wedge (Does(s,a,t1) \diamond\!\!\rightarrow \sim\phi)]$	$\forall$E, 3
18	$\quad \sim(Does(s,a,t1) \diamond\!\!\rightarrow \sim\phi)$	DS, 14, 17
19	$\quad Does(s,a,t1) \boxdot\!\!\rightarrow \phi$	L, 18
20	$\quad\quad Does(s,a,t1)$	
21	$\quad\quad \phi$	, 19, 20
22	$\quad\quad \phi \rightarrow \psi$	, 16, 20
23	$\quad\quad \psi$	$\Rightarrow$E, 21, 22
24	$\quad Does(s,a,t1) \boxdot\!\!\rightarrow \psi$	, 20–23
25	$\quad Does(s,a,t1) \diamond\!\!\rightarrow \sim\psi$	$\wedge$E, 10
26	$\quad \sim(Does(s,a,t1) \boxdot\!\!\rightarrow \psi)$	L, 25
27	$\quad \bot$	
28	$\psi \wedge \sim\exists x\exists\alpha\exists t'[Can(x,\alpha,t') \wedge (Does(x,\alpha,t') \diamond\!\!\rightarrow \sim\psi)]$	
29	$\mathbf{N}\psi$	

Bibliography

Anscombe, G. E. M., 1981. “Causality and Determination”, in *Metaphysics and the Philosophy of Mind*, Minneapolis: University of Minnesota Press.

Armstrong, D., 1983. *What is a Law of Nature?*, Cambridge: Cambridge University Press.

Armstrong, D., 1997. *A World of States of Affairs*, Cambridge: Cambridge University Press.

Austin, J. L., 1956. “Ifs and Cans”, *Proceedings of the British Academy* 42: 109-132.

Ayer, A. J., 1954. "Freedom and Necessity", in *Philosophical Essays*, London: Macmillan.

Backmann, M., 2016. "I Tensed the Laws and the Laws Won: Non-Eternalist Humeanism", *Manuscrito* 39 (4): 255-277.

Backmann, M., and Reutlinger, A., 2014. “Better best systems - Too good to be true”, *Dialectica* 68 (3): 375–390.

Bailey, A. M., 2012. “Incompatibilism and the Past”, *Philosophy and Phenomenological Research* 85: 351–376.

Beebee, H., 2000. “The Nongoverning Conception of Laws of Nature”, *Philosophy and Phenomenological Research* 61: 571–594.

Beebee, H., 2002. “Reply to Huemer on the Consequence Argument”, *The Philosophical Review* 111 (2): 235–241.

Beebee, H., 2003. “Local Miracle Compatibilism”, *Noûs* 37 (2): 258–277.

Beebee, H., and Mele A.R., 2002. "Humean compatibilism”, *Mind* 111: 201–23.

Bennett, J., 1984. "Counterfactuals and Temporal Direction", *The Philosophical Review* 93 (1): 57-91.

Berofsky, B., 2003. “Classical Compatibilism: Not Dead Yet”, in Widerker and McKenna 2003, 107–126.

Bigelow, J., 1999. "Scientific Ellisianism", in *Causation and Laws of Nature*, ed. Sankey, H., Kluwer Academic Publishers, 45–59.

Bird, A., 1998. “Dispositions and Antidotes”, *Philosophical Quarterly* 48: 227-34.

Bird, A., 2005. “The dispositionalist conception of laws”, *Foundations of Science* 10: 353-370.

Bird, A., 2007. *Nature’s Metaphysics: Laws and Properties*, Oxford: Oxford University Press.

Blum, A., 2000. “N”, *Analysis* 60 (3): 284-286.

Bonevac, D., 2003. *Deduction: Introductory Symbolic Logic*, 2nd edition. Oxford: Blackwell Publishing.

Bostock, S., 2001, *The Necessity of Natural Laws*, Ph.D. thesis, University of Sheffield.

Brueckner, A. 2008. “Retooling the Consequence Argument”, *Analysis* 68: 10–13.

Campbell, J. K., 2007. “Free Will and the Necessity of the Past”, *Analysis* 67 (2): 105–111.

Campbell, J. K., 2008. “Reply to Brueckner”, *Analysis* 68: 264–269.

Campbell, J. K., 2010. “Incompatibilism and Fatalism: Reply to Loss”, *Analysis* 70: 71–76.

Carlson, E., 2003. "On a New Argument for Incompatibilism", *Philosophia* 31 (1–2): 159–164.

Cartwright, N., 1983. *How the Laws of Physics Lie*, Oxford: Clarendon Press.

Cartwright, N., 1989. *Nature's Capacity and Their Measurement*, Oxford: Clarendon Press.

Cartwright, N., 2000. "Against the Completability of Science", in *The Proper Ambition of Science*, eds. J. Wolff, M.W.F. Stone, New York: Routledge.

Cartwright, N., 2009. "Causal laws, policy predictions and the need for genuine powers", in Dispositions and Causes, ed. Handfield, T., Oxford: Oxford University Press.

Cartwright, N., 2010. "Natural Laws and the Closure of Physics", in *Visions of Discovery. New Light on Physics, Cosmology and Consciousness*, ed. Chiao, R.Y., Leggett, A.J., Cohen, M.L. & Harper, C.L, Cambridge: Cambridge University Press, 612-622

Cartwright, N. & Pemberton, J., 2013. "Aristotelian Powers: Without Them, What Would Modern Science Do?", in *Powers and Capacities in Philosophy: The New Aristotelianism*, eds. Greco, J. & Gross, R., New York: Routledge, 93-112.

Chisholm, R., 1964. "Human Freedom and the Self", The Lindley Lecture, Reprinted in Watson 2003.

Chisholm, R., 1966. "Freedom and action", in *Freedom and determinism*, ed. Lehrer, K., New York: Random House.

Choi, S., 2006. "The Irrelevance of Thomas Crisp and Ted Warfield's Desiderata and Proposed Counterexamples to Principle Beta", *Philosophia Christi* 8 (2): 421-438.

Clarke, R., 2003. *Libertarian accounts of free will*, New York: Oxford University Press.

Clarke, R., forthcoming. "Free will, agent causation, and 'disappearing agents'", *Nous*: 1-21.

Cohen, J. & Callender, C., 2009. "A better best system account of lawhood", *Philosophical Studies* 145: 1-34.

Crisp, T. and Warfield, T., 2000. "The Irrelevance of Indeterministic Counterexamples to Principle Beta", *Philosophy and Phenomenological Research* 61 (1): 173-184.

Cutter, B., 2017. "What is the consequence argument an argument for?", *Analysis* 77 (2): 278:287.

Davidson, D., 1963. "Actions, Reasons, and Causes", *Journal of Philosophy* 60: 685–700.

Demarest, H., 2017. "Powerful properties, powerless laws", in *Putting Powers to Work: Causal Powers in Contemporary Metaphysics*, ed. Jacobs, J., Oxford: Oxford University Press.

Dorr, C., 2016. "Against Counterfactual Miracles", The Philosophical Review, 125 (2): 241–286.

Dretske, F. I., 1977. "Laws of Nature", *Philosophy of Science* 44 (2): 248-268.

Earman, J., 1978. "The universality of laws", *Philosophy of Science* 45 (2): 173-181.

Earman, J., 1986. *A Primer on Determinism*, Dordrecht Netherlands: Reidel.

Ekstrom, L. W., 1993. "A coherence theory of autonomy", *Philosophy and Phenomenological Research* 53: 599–616.

Ekstrom, L. W., 1998. "Freedom, causation and the consequence argument", *Synthese* 115 (3): 333-54

Ekstrom, L. W., 2000. *Free will: A philosophical study*, Boulder: Westview Press.

Ellis, B., 2001. *Scientific Essentialism*, Cambridge: Cambridge University Press.

Ellis, B., 2002. *The Philosophy of Nature*, Chesman: Acumen.

Ellis, B. and Lierse, C., 1994. "Dispositional essentialism", *Australasian Journal of Philosophy* 72: 27–45.

Fara, M., 2008. "Masked Abilities and Compatibilism", *Mind* 117: 843–865.

Finch, A., 2013a. "On Behalf of the Consequence Argument: Time, Modality, and the Nature of Free Action", *Philosophical Studies*, 163(1): 151–170.

Finch, A., 2013b. "Against Libertarianism", *Philosophical Studies* 166 (3): 475-493.

Finch, A. and Warfield, T., 1998. "The *Mind* argument and libertarianism", *Mind* 107: 515-528.

Fine, K., 1975. "Critical Notice of Lewis, Counterfactuals", *Mind* 84 (335): 451-458.

Fine, K., 2012. "Guide to Ground" in *Metaphysical Grounding: Understanding the Structure of Reality*, eds. Correia, F. and Schnieder, B., Cambridge: Cambridge University Press.

Fischer, J. M., 1988. "Freedom and Miracles", *Noûs* 22(2): 235–252.

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

Fischer, J. M., 1999. "Recent work on moral responsibility", *Ethics* 110: 91-139.

Fischer, J. M., and Tognazzini, N. A., 2013. "The Logic of Theological Incompatibilism: A Reply to Westphal", *Analysis* 73: 46–48.

Flint, T., 1987. "Compatibilism and the Argument From Unavoidability", *Journal of Philosophy* 84: 423-440.

Foley, R., 1979. "Compatibilism and Control Over the Past", *Analysis* 39: 70-74.

Frankfurt, H., 1969. "Alternate Possibilities and Moral Responsibility", *Journal of Philosophy* 66: 820-839.

Frankfurt, H., 2003. “Some Thoughts concerning PAP”, in Widerker and McKenna 2003.

Franklin, C. E., 2011. “Farewell to the luck (and mind) argument”, *Philosophical Studies* 156: 199–230.

Franklin, C. E., 2014. “Event-Causal Libertarianism, Functional Reduction, and the Disappearing Agent Argument”, *Philosophical Studies* 170: 413–432.

Gallois, A., 1977. “Van Inwagen on Free Will and Determinism”, *Philosophical Studies* 32 (1): 99–105.

Ginet, C., 1983. “In defence of incompatibilism”, *Philosophical Studies* 44: 391-400.

Ginet, C., 1990. *On Action*, Cambridge: Cambridge University Press.

Goldman, A., 1970. *A Theory of Human Action*, Englewood Cliffs, NJ: Prentice-Hall.

Goodman, J., 2015. “Knowledge, Counterfactuals and Determinism”, *Philosophical Studies* 172 (9): 2275-2278.

Gorham, G., 2004. "Cartesian Causation: Continuous, Instantaneous, Overdetermined", *Journal of the History of Philosophy* 42: 389–423

Graham, P. A., 2008. "A defense of local miracle compatibilism", *Philosophical Studies* 140: 65–82.

Graham, P. A., 2010. "Against the Mind Argument", *Philosophical Studies* 148: 273–294.

Griffith, M., 2010. "Why agent-caused actions are not lucky", *American Philosophical Quarterly* 47: 43–56.

Haji, Ishtiyaque, 2009. *Incompatibilism's Allure*, Buffalo: Broadview Press.

Handfield, T., 2001, "Dispositional essentialism and the possibility of a law-abiding miracle", *Philosophical Quarterly* 51: 484–494.

Handfield T., 2004. "Counterlegals and necessary laws", *Philosophical Quarterly* 54: 402-419.

Handfield, T. (ed.), 2009. *Dispositions and Causes*, Oxford: Clarendon Press.

Harré, R. and Madden, E. H., 1975. *Causal Powers: A Theory of Natural Necessity*, Oxford: Blackwell.

Hobart, R. E., 1934. "Free will as involving determinism and inconceivable without it", *Mind* 43: 1-27.

Huemer, M., 2000. "Van Inwagen's Consequence Argument", *The Philosophical Review* 109 (4): 525–544.

Huemer, M. and Kovitz, B., 2003. "Causation as Simultaneous and Continuous", *The Philosophical Quarterly* 53: 556–565.

Jackson, F., 1977. "A Causal Theory of Counterfactuals", *Australasian Journal of Philosophy* 55: 3-21

Jordan, Z., 1963. "Logical Determinism", N*otre Dame Journal of Formal Logic*, Volume IV, Number 1.

Kane, R., 1996. *The significance of free will*, New York: Oxford University Press.

Kane, R., 1999. "Responsibility, luck, and chance: Reflections on free will and indeterminism", *Journal of Philosophy* 96: 217–240.

Kane, R., 2011. *The Oxford Handbook of Free Will*, Second Edition, Oxford: Oxford University Press.

Kenny, A., 1975. *Will, Freedom, and Power*, Oxford: Blackwell.

Kistler, M., 2002. "The Causal Criterion of Reality and the Necessity of Laws of Nature", *Metaphysica* 3/1: 57–86.

Kripke, S., 1980. *Naming and Necessity*, Harvard: Harvard University Press.

Lehrer, K.. 1968. "Cans Without Ifs", *Analysis* 29 (1): 29–32.

Lehrer, K., 1976. "'Can' in Theory and Practice: A Possible Worlds Analysis", in *Action Theory*, eds. Miles Brand and D. Walton, Dordrecht: D. Reidel.

Levy, N. and McKenna, M., 2009. "Recent Work on Free Will and Moral Responsibility", *Philosophy Compass* 4 (1): 96-133.

Lewis, D., 1973. *Counterfactuals*, Cambridge: Harvard University Press.

Lewis, D., 1979. "Counterfactual Dependence and Time's Arrow", *Noûs* 13: 455–76

Lewis, D., 1981. "Are We Free to Break the Laws?", *Theoria* 47: 113 – 21

Lewis, D., 1983. "New work for a theory of universals", *Australasian Journal of Philosophy* 61: 343–377.

Lewis, D., 1986. *On the plurality of worlds*, Oxford: Basil Blackwell.

Lewis, D., 1994. "Humean supervenience debugged", *Mind* 103 (412): 473–490.

Lewis, D., 1997. "Finkish Dispositions", *The Philosophical Quarterly* 47: 143–158.

Loewer, B., 1996. "Humean supervenience", *Philosophical Topics* 24 (1): 101–127.

Loewer, B., 2007. "Laws and natural properties", *Philosophical Topics* 35 (1,2): 313–328

Loss, R., 2009. "Free Will and the Necessity of the Present", *Analysis* 69: 63–69.

Loss, R., 2010. "Fatalism and the Necessity of the Present: Reply to Campbell", *Analysis* 70: 76–78.

Lowe, E. J., 2001. "Event causation and agent causation", *Grazer Philosophische Studien* 61: 1-20.

Lowe, E. J., 2002. *A Survey of Metaphysics*, Oxford: Oxford University Press.

Lukasiewicz, J., 1970. “On Determinism”, in *Selected Works*, ed. Borkowski, L., North-Holland, Amsterdam, 110–128.

Martin, C.B., 1994. “Dispositions and Conditionals”, *The Philosophical Quarterly* 44: 1–8.

McKay, T. J. and Johnson, D. 1996. “A Reconsideration of an Argument against Compatibilism”, *Philosophical Topics* 24 (2): 113–122.

McKenna, M., 2001. "Source Incompatibilism, Ultimacy, and the Transfer of Non-Responsibility", *American Philosophical Quarterly* 38: 37-52.

McKenna, M., 2010. "Whose Argumentative Burden, Which Incompatibilist Arguments? – Getting the Dialectic Right", *Australasian Journal of Philosophy* 88: 429–443.

Mele, A. R., 2000. "Deciding to act", *Philosophical Studies* 100: 81-108.

Mele, A. R., 2004. "Discussion — Velleman on action and agency", *Philosophical Studies* 121: 249–261.

Mele, A. R., 2006. *Free will and luck*, New York: Oxford University Press.

Mele, A. R., 2014. "Luck and Free will," *Metaphilosophy* 45: 543:557.

Merlussi, P., 2016. "Determinism, laws of nature and the consequence argument", *Manuscrito* 39: 73-95.

Merlussi, P. & Cartwright, N., forthcoming. "Are laws of nature consistent with contingency?", in *Laws of nature*, eds. W. Ott and L. Patto, Oxford: Oxford University Press.

Molnar, G., 2003. *Powers: a study in metaphysics*, Oxford: Oxford University Press.

Montague, R., 1974. “Deterministic Theories,” in *Formal Philosophy*, ed. Thomason, R. H., New Haven, CT: Yale University Press: 303–359.

Moore, G. E., 1912. “Free Will” in his *Ethics*, Oxford: Oxford University Press, Chapter 6.

Mortari, C., 2011. *Lógica III*, Retrieved from https://ead.ufsc.br/filosofia/files/2011/04/logicaIII.pdf.

Müller, T., and Placek, T., 2015. “Defining determinism”, *The British Journal for the Philosophy of Science* 0: 1-38.

Narveson, J., 1977. “Compatibilism Defended”, *Philosophical Studies* 32: 83-87.

Nelkin, D., 2001. "The Consequence Argument and the *Mind* Argument", *Analysis* 61: 107-115.

Nowell-Smith, P. H., 1948. "Free will and moral responsibility", *Mind* 55: 45-61

Nowell-Smith, P. H., 1957. *Ethics*, Oxford: Blackwell Publishers.

Oakley, S., 2006. “Defending Lewis’s Local Miracle Compatibilism”, *Philosophical Studies* 130 (2) : 337–349.

O’Connor, T., 1993a. “On the transfer of necessity”, *Nous* 27: 204-18.

O’Connor, T., 1993b, “Indeterminism and free agency: three recent views”, *Philosophy and Phenomenological Research* 53: 499-526.

O’Connor, T., 2000. *Persons and causes: The metaphysics of free will*, New York: Oxford University Press.

O’Connor, T., 2009. "Agent-causal power", in *Dispositions and causes*, ed. Handfield, T., New York: Oxford University Press, 189–214.

O’Connor, T., 2011. "Agent-causal Theories of Freedom", in Kane 2011.

Pendergraft, G., 2010. "The explanatory power of local miracle compatibilism", *Philosophical Studies* 156: 249–66.

Pereboom, D., 2001. *Living Without Free Will*, Cambridge: Cambridge University Press.

Pereboom, D., 2007. "Hard incompatibilism", in *Four views on free will*, eds. Fischer, J. M. et al., Malden: Blackwell.

Pereboom, D., 2015. "The phenomenology of agency and deterministic agent causation", *Horizons of authenticity in phenomenology, existentialism, and moral psychology*, eds. Pedersen, H., Altman, M., 277–294.

Plantinga, A., 1974. *The Nature of Necessity*, Oxford: Oxford University Press.

Pruss, A. R. 2013. "Incompatibilism Proved", *Canadian Journal of Philosophy* 43: 430–437.

Quine, W. V. O., 1970. *Philosophy of logic*, Cambridge, MA: Harvard University Press.

Reichenbach, H., 1947. *Elements of Symbolic Logic*, New York: the Macmillan Company.

Roberts, J., 2006. "Determinism", in *The Philosophy of Science: An Encyclopedia*, eds. Pfeifer, J. and Sarkar, S., Routledge, New York, 197-208.

Runyan, J., D., 2016. "Events, agents ad settling whether and how one intervenes", *Philosophical Studies* 173: 1629–46

Sartorio, C., forthcoming. "Frankfurt-style examples", in *Routledge Companion to Free Will*, eds. Timpe, K., Griffith, M., Levy, N., New York: Routledge.

Sehon, S., 2011. "A flawed conception of determinism in the consequence argument", *Analysis* 71: 30-38.

Shabo, S., 2013. "Free will and mystery: looking past the mind argument," *Philosophical Studies* 162: 291–307.

Shalkowski, S., 2004. "Logic and Absolute Necessity", *Journal of Philosophy* CI 2: 55-82.

Shoemaker, S., 1980. "Causality and properties", in *Time and Cause*, ed. van Inwagen, P., Dordrecht: Reidel, 109-135.

Shoemaker, S., 2003. "Caring, identification, and agency", *Ethics* 114: 88–118.

Sider, T., 2010. *Logic for Philosophy*, New York: Oxford University Press.

Sinnott-Armstrong, W., 1999. "Begging the question", *Australasian Journal of Philosophy* 77 (2): 174-191.

Slote, M., 1982. "Selective Necessity and the Free Will Problem", *Journal of Philosophy* 79 (1): 5–24.

Stalnaker, R., 1968. "A Theory of Conditionals", in *Studies in Logical Theory*, ed. Rescher, N., Oxford: Blackwell.

Stalnaker, R., 1984. *Inquiry*, Cambridge, MA: Bradford Books.

Steward, H., 2012. *A metaphysics for freedom*, Oxford: Oxford University Press.

Swoyer, C., 1982. "The nature of natural law", *Australasian journal of philosophy* 60: 203-23.

Taylor, B., 1993. "On natural properties in metaphysics", *Mind* 102: 81–100.

Tooley, M., 1987. *Causation: A Realist Approach*, Oxford: Clarendon Press.

van Fraassen, B., 1980. *The Scientific Image*, Oxford: Clarendon Press.

van Fraassen, B. C., 1989. *Laws and symmetry*, Oxford: Clarendon.

Van Inwagen, P., 1977a. "Reply to Narveson", *Philosophical Studies* 32: 89-98.

Van Inwagen, P., 1977b. "Reply to Gallois", *Philosophical Studies* 32: 107-111.

van Inwagen, P., 1983. *An Essay on Free Will*, Oxford: Clarendon Press.

van Inwagen, P., 2000. "Free Will Remains a Mystery", in *The Oxford Handbook of Free Will*, ed. Kane, R., Oxford: Oxford University Press, 158-177.

van Inwagen, P., 2004. "Freedom to Break the Laws", *Midwest Studies in Philosophy* 28: 334–350

van Miltenburg, N. and Ometto, D., forthcoming. "The libertarian predicament: a plea for action theory", *Synthese*: 1-18.

Vessel, J. P., 2003. "Counterfactuals for Consequentialists", *Philosophical Studies* 112 (2): 103–125.

Vihvelin, K., 2004. "Free Will Demystified: A Dispositional Account", *Philosophical Topics* 32 (1–2): 427–450.

Vihvelin, K., 2013. *Laws, causes and free will: why determinism doesn't matter*, Oxford: Oxford University Press.

Vihvelin, K., 2015. "How not to think about free will", *Journal of Cognition and Neuroethics* 3 (1): 393: 403.

Vihvelin, K., 2017. "Arguments for incompatibilism", Retrieved from: https://plato.stanford.edu/entries/incompatibilism-arguments/.

Warfield, T., 2000. "Causal determinism and human freedom are incompatible: a new argument for incompatibilism", *Philosophical Perspectives* 14: 167–80.

Warfield, T., 2003. "Compatibilism and Incompatibilism: Some Arguments", in *The Oxford Handbook of Metaphysics*, ed. Michael Louxand, Dean Zimmerman, Oxford: Oxford University Press, 613–630.

Watson, G., 1975. "Free agency", *Journal of Philosophy* 72: 205–220.

Watson, G., 2003. *Free Will*, Second Edition, Oxford: Oxford University Press.

Westphal, J., 2012. "Is There a Modal Fallacy in van Inwagen's 'First Formal Argument'?", *Analysis* 72 (1): 36–41.

Wheeler, B., 2015. "Simplicity, Language-Dependency and the Best System Account of Laws", *Theoria. An International Journal for Theory, History and Foundations of Science* 31 (2): 189-206.

White, M. J., 1984. "The Necessity of the Past and Modal-Tense Logic Incompleteness", *Notre Dame Journal of Formal Logic* (1): 59–71.

Widerker, D., 1987. “On an Argument for Incompatibilism” *Analysis* 47 (1): 37–41.

Widerker, D. and McKenna, M. eds., 2003. *Moral Responsibility and Alternative Possibilities*, Burlington, VT: Ashgate.

Wierenga, E., 2011. "Tilting at Molinism", in *Molinism: The contemporary debate*, ed. Perszyk, K., Oxford: Oxford University Press, 118–139.

Wiertz, O., 2015. “Is Plantinga’s A/c Model an Example of Ideologically Tainted Philosophy?”, in *Plantinga’s ‘Warranted Christian Belief’: Critical Essays with a Reply by Alvin Plantinga*, ed. Schönecker, D., De Gruyter, 83-114.

Wiggins, D., 1973. "Towards a reasonable libertarianism", in *Essays on freedom and action*, ed. Honderich, T., London: Routlegde and Keegan Paul.

Wimsatt, W., 1992. "Golden Generalities and Co-opted Anomalies: Haldane vs. Muller and the Drosophila Group on the Theory and Practice of Linkage Mapping" in *Fisher, Haldane, Muller, and Wright: Founders of the Modern Mathematical Theory of Evolution*, ed. Sarkar, S., Dordrecht: Martinus Nijhoff, 107-166.