What ‘If’? A Modal Analysis of Indicative Conditionals

Finlay McCardel MA (Hons) MSc Glas

Submitted in fulfilment of the requirements for the
Degree of Doctor of Philosophy

School of Humanities
College of Arts
University of Glasgow

September 2023

Typeset in LaTeX.

Abstract

Indicative conditionals - e.g. 'If I was born in Glasgow, then I was born in Scotland' - seem to express propositions. In other words, they seem to express thoughts that can be shared and evaluated as either true or false. For the past hundred years or so, analytic philosophers have commonly interpreted them as equivalent to disjunctions, e.g. 'Either I was not born in Glasgow, or I was born in Scotland'. Those who dig a little deeper tend to agree that this is not quite right, but they often come to the conclusion that indicative conditionals do not express propositions at all. I argue for an alternative theory according to which indicative conditionals express a type of strict implication. What this means is that they are modal statements - e.g. '*Necessarily*, either I was not born in Glasgow, or I was born in Scotland' - an idea that used to be quite popular despite striking many contemporary philosophers as radical. More specifically, I defend the idea that indicative conditionals express *metaphysical* necessity. I argue that this view respects our intuitions in an important way, and I show that it can be incorporated into a promising view of natural language conditionals more generally. Lastly, I argue that it also has plausible consequences for the related concepts of conditional probability, causal explanation, and risk.

Contents

Acknowledgements

I am extremely grateful to the SGSAH AHRC DTP Executive Committee for granting me the Collaborative Doctoral Award that made completing this doctoral thesis a practical possibility. I am also extremely grateful to my wonderful dad, Tom, for the extra support that kept that possibility alive a little longer.

Special thanks are owed to C. Michael Holloway (my industry partner at NASA Langley Research Center) and Neil McDonnell, J. Adam Carter, and Adam Rieger (my supervisors at the University of Glasgow), all of whom were crucial links in the causal chain that led to these words being written.

To my friends and philosophical peers, Matthew Kinakin, Laura Fearnley, and Bazil Hughes: thank you for sharing your thoughts with me, and for showing an interest in mine. To my friends and philosophical mentors, Brian King, Tim Button, and Ben Colburn: thank you for sharing your wisdom. And to my gorgeous boyfriend, Francisco, with whom I want to share everything: thank you for being so patient. I dedicate this thesis to you.

Declaration

To the best of my knowledge, the material presented in this thesis is original (except where acknowledged in the usual way) and has not previously been submitted in whole or part for a degree at any university.

Finlay McCardel MA (Hons) MSc Glas

Chapter 1

Introduction

Philosophers sometimes like to think about logical connectives, expressed in English by words like ‘and’, ‘or’, and ‘if’. More specifically, they like to think about the way that each logical connective influences the *truth conditions* of a proposition, i.e. the conditions under which it is true. Some logical connectives influence truth conditions in extremely simple ways. For example, the truth value (i.e. the *truth* or *falsehood*) of the conjunction expressed by ‘Finn is an uncle and Maia is Finn’s niece’ is simply a function of the truth values of its constituent propositions: if it is true that Finn is an uncle, and it is true that Maia is Finn’s niece, then it is true that Finn is an uncle and Maia is Finn’s niece; otherwise, it is false. In philosophical parlance, the conjunction is *truth-functional*.

This thesis is about conditionals: a type of compound proposition which is typically expressed in English using the word ‘if’. For example: ‘Finn is an uncle *if* Maia is Finn’s niece.’ This *indicative* conditional (named this way because of the grammatical mood in which it is expressed) can be represented as ‘C if A’, where C is the conditional’s consequent (i.e. the proposition expressed by ‘Finn is an uncle’) and A is the conditional’s antecedent (i.e. the proposition expressed by ‘Maia is Finn’s niece’). The letters ‘C’ and ‘A’ are variables, meaning they can represent different propositions in different contexts; however, they always represent the consequent and antecedent respectively. Switching the order of a conditional’s constituent propositions without making any other changes gives us a different conditional: in this case, ‘Maia is Finn’s niece *if* Finn is an uncle’, which is much less plausible than the original conditional. In other words, the operation performed by ‘if’ is not commutative. However, we *can* reverse the order of C and A if we also move the word ‘if’ to the beginning of the sentence, using a comma (and possibly the word ‘then’) to separate the constituent parts, as in ‘*If* Maia is Finn’s niece, *then* Finn is an uncle.’

Much of the philosophical literature on conditionals is devoted to the question of whether they are truth-functional. That is: Does ‘If A, then C’ express a proposition whose truth value depends solely on the truth values of A and C? It is quite common nowadays for analytic philosophers to hold views according to which the answer to this question is 'yes' - see e.g. Whitehead and Russell (1910), Grice (1967a), Jackson (1979), Rieger (2013), and Williamson (2020). However, I think the answer is 'no'. My answer may seem at first to make 'if' an outlier, since 'and' and 'or' both intuitively yield truth-functional compound propositions. Yet there are other connectives in our language that intuitively do not yield truth-functional compound propositions. Consider e.g. 'Maia is Finn's niece *because* Finn is an uncle' - the truth value of this statement is not a simple function of the truth values of its constituent parts. Rather, the word 'because' seems to imply the presence of some kind of connection between the constituent propositions. In this regard, 'if' may be more like 'because' than like 'and'.

I argue that a conditional is true iff[1] there is a particular kind of *modal* connection between the antecedent and consequent: namely, a connection of metaphysically strict implication. What this means is that 'If A, then C' is true iff it is metaphysically[2] impossible that A is true and C is false. Equivalently: 'If A, then C' is true iff it is metaphysically necessary that either A is false or C is true. This is not a new idea. In fact, the idea that conditionals express some sort of strict implication used to be commonplace - see e.g. MacColl (1880), Peirce (1896), and especially C. I. Lewis (1912, 1914, 1918). Nowadays, however, it is usually dismissed out of hand, ignored, or forgotten.

By filling in the details in my own particular way, I hope to make some progress towards reviving the idea that indicative conditionals are strict. Some of those details are of interest in and of themselves. For example, in chapter 6, I defend the idea that natural language involves just *one* logical type of conditional. The result is a view according to which most of the conditionals used in daily life are false, including (no doubt) some of those used in this thesis. However, I argue (in chapter 4) that we may assert certain false conditionals under the right conditions. In contemporary analytic philosophy, these are controversial ideas.

It may not be immediately apparent why it is important to have a good philosophical theory of conditionals, since the word 'if' may seem insignificant. Yet it is used to express a vital type of thought. As Dorothy Edgington writes:

> The ability to think conditional thoughts is a basic part of our mental equipment. A view of the world would be an idle, ineffectual affair without them.

[1] I use 'iff' throughout this thesis to mean the biconditional connective 'if and only if' (as is standard). This may seem viciously circular, as though conditionals are being analysed in terms of more conditionals. That is not so, since 'iff' does not feature in the analysans - rather, it relates the analysans to the analysandum. Assuming that biconditionals are the correct way to express philosophical analyses, a theory of conditionals has an impact on what it means to provide a philosophical analysis. The theory of conditionals defended in this thesis implies that 'iff' indicates necessary and sufficient conditions (as is standard); in particular, it implies that 'iff' indicates *metaphysically* necessary and sufficient conditions.

[2] An elaboration on 'metaphysically' will have to wait until chapter 3.

> There's not much point in recognising that there's a predator in your path unless you also realise that if you don't change direction pretty quickly you will be eaten. (Edgington, 1995, p. 235)

Given that conditional thoughts are such an integral part of the way we view the world, it is quite surprising that there is still so much disagreement about them, even among the experts. Is there a principled way of deciding which conditional thoughts we should have? If so, in what sense *should* we have them? Does it depend on our subjective view of the world, or are some conditionals objectively good and others objectively bad? If the latter, are these conditionals true and false respectively, or are they *non-truth-evaluable*? Philosophers are yet to agree on the answers to these questions.

As a demonstration of their significance, conditionals also feature in the philosophical analyses of widely used concepts such as *causation*. This makes it even more important to make good sense of them. As Stalnaker writes:

> Small differences among analyses of conditionals may have consequences for many complex constructions involving conditionals. A small distortion in the analysis of the conditional may create spurious problems with the analysis of other concepts. (Stalnaker, 1981, p. 87)

In other words, conditionals are not isolated: if we cannot fully understand them, then we cannot fully understand certain other concepts either. With this in mind, I argue (in chapter 7) that my view of conditionals can help us to make sense of the intuitive connection between conditionals and *conditional probability*. In particular, I argue that a conditional 'If A, then C' is true iff the conditional probability of C given A is 1. This connection is intuitively plausible by itself, but it is also instrumental in clarifying other philosophical debates, so the fact that my view of conditionals is consistent with it is a big part of the argument in favour of my view. In chapter 8, I show that this connection can help us to make sense of *causal explanation*. Finally, in chapter 9, I use the ideas discussed in chapter 7 to clarify a recent philosophical debate regarding the nature of *risk*.

By writing this thesis, I do not mean to suggest that we must straighten out our theory of conditionals before we can get on with using them - apart from anything else, that would be self-defeating, since I make extensive use of them in this thesis. As with most things, a rough understanding usually gets the job done. Still, it would be nice to have a clearer understanding, even if only for its own sake.

Chapter 2

The Literature On Conditionals

The word ‘if’ has generated a great deal of philosophical literature for something so seemingly unremarkable. More than that, for something so seemingly *simple*, it has generated a great deal of *complex* philosophical literature. In this chapter, I give an overview of some of the most influential theories of conditionals in analytic philosophy, highlighting some of their consequences and explaining some key concepts along the way. Doing so will make it easier for the reader to understand and evaluate the theory that I go on to defend.

I begin by explaining the material interpretation of indicative conditionals, and the distinction between indicative conditionals and counterfactual conditionals. I then explain David Lewis's influential theory of counterfactual conditionals and the conceptual scheme of *possible worlds* that comes with it. In section 2.3, I explain the concept of *strict implication*, and I point to some early endorsements of the idea that conditionals can be analysed in terms of it. I then distinguish strict conditionals from *variably* strict conditionals, using Stalnaker's theory of conditionals to illustrate the latter. In section 2.5, I look at some examples of unified theories, i.e. theories that apply the same analysis to both indicative and counterfactual conditionals. And in section 2.6, I explain why some contemporary philosophers have been driven to the conclusion that indicative conditionals are *non-truth-evaluable*.

My aim is not to provide a comprehensive list of the many arguments that have been made for and against the different theories of conditionals that exist, nor to provide a comprehensive list of the theories themselves. Rather, my aim is to hone in on those theories and arguments that are especially relevant to the content of subsequent chapters. In that vein, I finish with some concluding remarks about what I hope the reader will take from all this: namely, a reason to continue reading.

2.1 The Material Interpretation

The early 20th century was a formative time for the philosophical literature on conditionals. One of the most influential philosophical texts from this period is Whitehead and Russell's *Principia Mathematica*, published in three volumes between 1910 and 1913. In the opening pages of the first volume, the authors list some logical symbols, stipulating their natural language interpretations. This includes a symbol for a type of *implication*: 'The symbol employed for "p implies q," *i.e.* for "$\sim p \vee q$," is "$p \supset q$." This symbol may also be read "if p, then q."' (Whitehead & Russell, 1910, p. 7) Here, '$\sim p \vee q$' may be read 'Either not-p or q'. Replacing 'p' and 'q' with 'A' and 'C' respectively gives us the following truth table for $A \supset C$:

A	C	$A \supset C$
T	T	T
T	F	F
F	T	T
F	F	T

$A \supset C$ is known as a 'material conditional', and the type of implication it expresses is known as:

material implication A materially implies C iff $A \supset C$ is true.

As can be seen from the truth table above, a material conditional $A \supset C$ is equivalent not just to 'Either not-A or C', but also to the negation of 'A and not-C'. In symbols:

material equivalence thesis $A \supset C \equiv \neg A \vee C \equiv \neg(A \wedge \neg C)$.[1]

The interpretation of 'If A, then C' as a material conditional is the only plausible truth-functional interpretation.[2] And it is *very* plausible. It has received defences from many philosophical heavyweights, including Whitehead and Russell, but also Grice (1967a), Jackson (1979), and Williamson (2020). In fact, the arguments in favour of it have been so successful that contemporary students of logic are commonly taught that it is true - see e.g. Goldfarb (2003).

Why has the material interpretation been so successful? One reason is that it often predicts the intuitively correct truth value for a conditional. Consider, for example:

[1] The material equivalence thesis assumes the standard *inclusive* interpretation of 'or', according to which 'A or C' is consistent with 'A *and* C'. It also assumes the standard *extensional* (i.e. non-modal) interpretation, according to which 'A or C' is true if at least one of its constituent propositions is true. An *intensional* (i.e. modal) interpretation will be explained in section 2.3. The material equivalence thesis also assumes the standard truth-functional interpretations of 'and' and 'not'.

[2] For an explanation of why this is, see Edgington (2020).

(I) *If Tolstoy did not write Anna Karenina, then someone else did.*

This conditional seems to be true. And that is exactly what the material interpretation predicts, because the proposition expressed by 'Tolstoy did not write Anna Karenina' is false, and all material conditionals with false antecedents are true (as can be seen from the truth table above). The material interpretation also makes sense of the fact that (I) seems to be interchangeable with the disjunction, 'Either Tolstoy wrote Anna Karenina or someone else did', as well as the negated conjunction, 'It is not the case that Tolstoy did not write Anna Karenina and nor did anyone else.' If (I) is material, then it is equivalent to both of these statements.

To be more precise, the material interpretation is not normally extended to *all* conditionals. Philosophers usually take it for granted that there are two different types of conditional, called 'indicative' (in reference to grammatical mood) on the one hand, and 'subjunctive' or 'counterfactual' (in reference to a tendency to be about unactualised situations) on the other. The material interpretation is restricted to the indicative category to which (I) belongs, whereas the counterfactual category is generally seen to require a more complicated, *modal* analysis. For an example of a counterfactual conditional, consider the following:

(II) *If Tolstoy* had *not written Anna Karenina, then someone else* would *have.*

This seems to be a different type of conditional to (I). After all, (II) does not seem to be interchangeable with the disjunction, 'Either Tolstoy wrote Anna Karenina or someone else did', nor the negated conjunction, 'It is not the case that Tolstoy did not write Anna Karenina and nor did anyone else.' Whilst both the disjunction and the negated conjunction seem to be true, (II) seems to be false.

Indicative conditionals and counterfactual conditionals are natural language constructions used by laypeople and philosophers alike. Let us say that, together, they comprise the category of *natural language conditionals*. When specificity does not matter, I write '>' to represent any natural language conditional. When it does matter, I write '$\rightarrow$' to represent an indicative conditional, and '$\Box\!\!\rightarrow$' to represent a counterfactual conditional (as per convention). The material interpretation of indicative conditionals can thus be summed up as follows: $\rightarrow$ is $\supset$.

There are many compelling arguments in favour of the material interpretation of indicative conditionals (e.g. those listed in Rieger (2013)), but as above, my aim is to cover only those arguments that are most relevant to subsequent chapters. In that vein, let us turn now to some arguments *against* the material interpretation. Consider the following:

G $\rightarrow$ E *If I was born in Glasgow, then I was born in England.*

Intuitively, **G** → **E** is false, since Glasgow is not in England. Yet according to the material interpretation, **G** → **E** is true, because all material conditionals with false antecedents are true, and (unfortunately) I was not born in Glasgow. Consider also:

L → **S** *If I was born in London, then I was born in Scotland.*

Intuitively, **L** → **S** is false, since London is not in Scotland. But according to the material interpretation, **L** → **S** is true, because all material conditionals with true consequents are true, and (fortunately) I was born in Scotland.

Examples such as these motivate what Jackson describes as the 'oldest and simplest objection to the material account', i.e. that 'it makes it too easy for a conditional to be true.' (1991, p. 2) This objection is a compelling one, but if it were the only one, I would not be looking for an alternative theory, in part because this objection can be responded to (as we will see in chapter 4). To my mind, a much more compelling objection can be found in the literature. Consider the following set of conditionals, which has been adapted from Ellis (1978, p. 118):

G → **S** *If I was born in Glasgow, then I was born in Scotland.*

G → **E** *If I was born in Glasgow, then I was born in England.*

L → **S** *If I was born in London, then I was born in Scotland.*

Those of us who know the locations of Glasgow and London are likely to find the following claim intuitive: **G** → **S** is true and **G** → **E** and **L** → **S** are both false. But the material interpretation of indicative conditionals is inconsistent with this claim. To see why, recall that a material conditional $A \supset C$ is equivalent to the disjunction $\neg A \vee C$. Hence, if **G** ⊃ **S** is true, then either I was not born in Glasgow or I was born in Scotland. But if I was not born in Glasgow, then **G** ⊃ **E** is true, because its antecedent is false; and if I was born in Scotland, then **L** ⊃ **S** is true, because its consequent is true.

The falsehood of the intuitive claim that **G** → **S** is true and **G** → **E** and **L** → **S** are both false is a surprising result, but to my mind, what makes this Ellisian paradox so compelling is the often overlooked fact that the proponent of the material interpretation must say that the intuitive claim is not only false, but *logically impossible*. After all, the material interpretation is an attempt to identify some of the laws of logic, and a proposition that is false according to the laws of logic is thereby logically impossible. Put simply: on the material interpretation, the claim that **G** → **S** is true and **G** → **E** and **L** → **S** are both false is *akin to contradiction*.

This is an alarming result. It is one thing to be told that something intuitively false is true, or that something intuitively true is false, but it is quite another to be told that something intuitively true is logically impossible. I think this warrants the search for an alternative theory to the material interpretation. Of course, humans do sometimes believe things that are logically impossible, but we are also sometimes told things that are false.

Note that the problem is not specific to the above set of conditionals, nor even to some interesting subset of indicative conditionals. Suppose, for instance, that the Portuguese and Qatari national football teams are competing against each other, and that the rules of the game dictate that they must keep playing until a winner is established. Now suppose that the game has come to an end, and consider the following conditionals:

$\neg\mathbf{P} \rightarrow \mathbf{Q}$	*If Portugal did not win the game, then Qatar won it.*
$\neg\mathbf{P} \rightarrow \neg\mathbf{Q}$	*If Portugal did not win the game, then Qatar did not win it (either).*
$\mathbf{P} \rightarrow \mathbf{Q}$	*If Portugal won the game, then Qatar won it (too).*

Besides being in the indicative mood, these conditionals do not seem to have anything in common with the conditionals above. Still, they can be used to illustrate the same point. Intuitively, $\neg\mathbf{P} \rightarrow \mathbf{Q}$ is true and $\neg\mathbf{P} \rightarrow \neg\mathbf{Q}$ and $\mathbf{P} \rightarrow \mathbf{Q}$ are both false. But the material interpretation makes this intuitive claim a logical impossibility: on the material interpretation, $\neg\mathbf{P} \rightarrow \mathbf{Q}$ is true iff its antecedent is false or its consequent is true; but in that case, either $\neg\mathbf{P} \rightarrow \neg\mathbf{Q}$ or $\mathbf{P} \rightarrow \mathbf{Q}$ must also be true, because the former has the same antecedent and the latter has the same consequent.

In fact, we do not need to reason from the truth of $\neg\mathbf{P} \rightarrow \mathbf{Q}$ in order to show that at least one of $\neg\mathbf{P} \rightarrow \neg\mathbf{Q}$ and $\mathbf{P} \rightarrow \mathbf{Q}$ must be true on the material interpretation. Without knowing anything about the content of the propositions **P** and **Q**, we can deduce that at least one of $\neg\mathbf{P} \supset \neg\mathbf{Q}$ and $\mathbf{P} \supset \mathbf{Q}$ must be true, since at least one of them must have a false antecedent or a true consequent.

I do not deny that the material conditional is a good model of the indicative conditional. Assuming the validity of *modus ponens* (i.e. the inference from A and $A \rightarrow C$ to C), an indicative conditional is true *only if* its antecedent is false or its consequent is true. This means that $A \supset C$ entails nothing that $A \rightarrow C$ does not also entail. For this reason, one can use material conditionals to model indicative conditionals that feature as premises in deductive arguments, safe in the knowledge that doing so will not give false justification to the argument's conclusion. The material conditional is also extremely simple. These characteristics make ⊃ a good model for →. But with the Ellisian paradox in mind, I think we should be wary of accepting the claim that → *is* ⊃.

Before moving on from the material interpretation, there is one more thing worth mentioning. In the literature on conditionals, while a good deal of attention has been given to the material interpretation's unintuitive consequences regarding the truth values of individual conditionals with false antecedents, an often overlooked fact is that there is also potential for conflict with intuition when it comes to the truth values of individual conditionals with *true* antecedents. Suppose we have someone who is known to have cheated at chess in the past - call them Niemann. Regarding a particular, current accusation of cheating, Niemann says that he is innocent. Now consider the following:

N → **T** *If Niemann says that he is innocent, then it is true that he is innocent.*

If it turns out that Niemann *is* innocent on this occasion, then the material interpretation renders **N** → **T** true, since its antecedent and consequent are both true. But there is at least *some* intuitive pull towards thinking that **N** → **T** is false, because it seems to posit a highly dubious *connection* between the conditional's antecedent and consequent. Counterexamples like this may invoke weaker intuitions than counterexamples with false antecedents, but they still deserve some attention, because some views have more to say about them than others. In particular, if conditionals express the kind of strict implication that I say they do, then **N** → **T** is probably false.

To make good sense of that strict implication, we first need to cover some more philosophical ground. In that vein, I turn next to Lewis's analysis of counterfactual conditionals, highlighting some of its intuitive and unintuitive consequences.

2.2 The Lewisian Analysis

David Lewis's classic text *Counterfactuals* (1973) presents what is now the most renowned philosophical theory of counterfactual conditionals. In it, Lewis makes extensive use of the concept of a *possible world*, which we may think of simply as a way that the actual world could be. One of the ways that the actual world could be is just the way that it is - in other words, the actual world is a possible world. But there are also many ways that the actual world could be that differ from how it is.

This conceptual scheme of possible worlds goes back at least as far as Leibniz, but Lewis is perhaps its main champion. According to Lewis (1986a), these other possible worlds should be understood as *metaphysical realities*: there really are other worlds, completely causally isolated from each other, each one just as real as the actual world - each one *being* the actual world from the perspective of its inhabitants. This is an ontological view, meaning it is a view about what exists. It is known by Lewis as 'modal realism', but by others (e.g. Stalnaker) as 'extreme realism' or 'extreme modal realism'. It is not a very popular view:

> Perhaps the biggest — if not the most philosophically sophisticated — challenge to Lewis's theory is "the incredulous stare", i.e., less colorfully put, the fact that its ontology is wildly at variance with common sense. (Menzel, 2023b, §2.1.5)

Yet it is undeniably very useful to speak *as though* other possible worlds exist. As Lewis writes:

> [P]hilosophers have offered a great many... analyses that make reference to possible worlds, or to possible individuals that inhabit possible worlds. I find that record most impressive. I think it is clear that... [it] has clarified questions in many parts of the philosophy of logic, of mind, of language, and of science - not to mention metaphysics itself. Even those who officially scoff often cannot resist the temptation to help themselves abashedly to this useful way of speaking. (D. K. Lewis, 1986a, p. 3)

Throughout this thesis, I speak as though other possible worlds exist, because doing so allows for greater clarity. I even draw inferences based on what these other possible worlds *would* be like if they *were* to exist. I effectively help myself to Lewis's conceptual scheme of possible worlds. I do so somewhat abashedly, not because I officially scoff at the idea that other possible worlds exist, but because I avoid the question entirely. I admit that this places a burden on me to do something that I do not do: namely, explain the usefulness of this way of speaking. However, just as it would be a mistake to let doubts about the ontological status of numbers stop us from doing maths, so it would be a mistake to let doubts about the ontological status of other possible worlds stop us from doing modal logic.

Here is how Lewis makes use of the conceptual scheme of possible worlds. Think of all possible worlds as being ordered in terms of their overall comparative similarity to the actual world. With this ordering in mind, we can express Lewis's analysis of counterfactual conditionals as follows: $A \mathrel{\Box\!\!\rightarrow} C$ is true iff C is true at all the closest (i.e. most similar) worlds at which A is true.[3] More accurately: $A \mathrel{\Box\!\!\rightarrow} C$ is true iff *either* there is a possible world at which both A and C are true that is closer to the actual world than any world at which A is true and C is false *or* there are no possible worlds at which A is true.

It may not be immediately apparent, but Lewis's analysis captures an intuitive idea about counterfactual conditionals, one which he expresses as follows:

> "*If kangaroos had no tails, they would topple over*' seems to me to mean something like this: in any possible state of affairs in which kangaroos have no tails, and which resembles our actual state of affairs as much as kangaroos having no tails permits it to, the kangaroos topple over.' (D. K. Lewis, 1973, p. 1)

[3] It may be that there is just one closest possible world in which A is true, but as we will see, it may also be that there are *ties* for the title of closest possible world.

Thus, the intuitive idea is that whether a counterfactual conditional is true depends on what happens in a world where (i) the conditional's antecedent is true, and (ii) the rest of the world is as similar to the actual world as the truth of the antecedent permits.

This Lewisian analysis of counterfactual conditionals is importantly different to the material interpretation of indicative conditionals. Whilst the truth values of material conditionals are determined solely by what goes on in the actual world, the truth values of counterfactual conditionals with false antecedents are determined, according to Lewis, by what goes on in *counterfactual* worlds. If we want to know the truth value of a material conditional, we may find it by investigating the actual world and discovering the truth value of its antecedent (and, if necessary, its consequent); but if we want to know the truth value of a counterfactual conditional with a false antecedent, we cannot go and investigate the relevant counterfactual world(s). Rather, we must rely on our intuitions. For this reason, on Lewis's view, when we consider a counterfactual conditional with a false antecedent, we generally have no choice but to assume that its intuitive truth value is vindicated by the relevant counterfactual world(s).

Sometimes, however, the antecedent of a counterfactual conditional turns out to be true (making 'counterfactual' a misnomer). In these cases, if the Lewisian analysis is right, $A \mathrel{\Box\!\!\rightarrow} C$ behaves just like a material conditional: it is true iff C is true. This makes sense when we consider that the closest possible A-world to the actual world when A is true is just the actual world itself.[4] With this in mind, it is no surprise that we can construct an analogous counterexample involving Niemann. Suppose once more that Niemann is known to have cheated at chess in the past, and as before, in response to a current accusation of cheating, Niemann says that he is innocent. But suppose this time that I assert the following conditional while (erroneously) believing that Niemann does not deny the accusation:

> $\mathbf{N} \mathrel{\Box\!\!\rightarrow} \mathbf{T}$ *If Niemann* were *to say that he is innocent, then it* would *be true that he is innocent.*

If it turns out that Niemann *is* innocent on this occasion, then the Lewisian analysis renders $\mathbf{N} \mathrel{\Box\!\!\rightarrow} \mathbf{T}$ true, because in that case, **T** is true at the closest possible **N**-world: namely, the world we are supposing to be actual. Yet there is at least some intuitive pull towards thinking that $\mathbf{N} \mathrel{\Box\!\!\rightarrow} \mathbf{T}$ is false - like its corresponding indicative conditional, it seems to posit a highly dubious *connection* between the conditional's antecedent and consequent. And just as some views have more than others to say about counterexamples like $\mathbf{N} \rightarrow \mathbf{T}$, so some views have more than others to say about counterexamples like $\mathbf{N} \mathrel{\Box\!\!\rightarrow} \mathbf{T}$. Even though my primary concern in this thesis is to defend a modal analysis of indicative conditionals, I argue (in chapter 6) that that modal analysis can

[4] This assumes that no possible world is as similar to the actual world as the actual world is to itself. I will not be questioning this assumption here.

be extended to counterfactual conditionals as well, yielding a *unified* analysis of natural language conditionals in general. If natural language conditionals express the kind of strict implication that I say they do, then $\mathbf{N} \mathrel{\Box\!\!\rightarrow} \mathbf{T}$ is probably false.

The above gives us an example of a relevant consequence of Lewis's analysis regarding the truth value of a counterfactual conditional considered in isolation. However, there are also some relevant consequences of Lewis's analysis regarding the truth values of counterfactual conditionals considered in combination with each other. Most of these consequences arise from Lewis's notion of *overall comparative similarity* between worlds. In particular, on Lewis's view, there may be *ties* in similarity between worlds.

The possibility of ties in similarity is something Lewis considers to be a virtue of his analysis. To see why, consider the following pair of conditionals, made famous by Lewis's PhD supervisor, Willard Van Orman Quine (1950, p. 15):

$\mathbf{C} \mathrel{\Box\!\!\rightarrow} \mathbf{I}$ *If Bizet and Verdi had been compatriots, Bizet would have been Italian.*

$\mathbf{C} \mathrel{\Box\!\!\rightarrow} \mathbf{F}$ *If Bizet and Verdi had been compatriots, Verdi would have been French.*[5]

One may well have the intuition that if Bizet and Verdi had been compatriots, then either Bizet would have been Italian or Verdi would have been French (since Verdi was Italian and Bizet was French). At the same time, it seems implausible that exactly one of the above conditionals is true, since there is nothing to decide between them, and it also seems implausible that they are *both* true, since they are contrary to each other. As Lewis (1973, p. 80) points out, his analysis can explain all this. According to Lewis, the set of closest worlds in which Bizet and Verdi are compatriots consists of worlds in which Bizet is Italian *and* worlds in which Verdi is French. Hence, even though the disjunction $\mathbf{I} \vee \mathbf{F}$ is true at all the closest $\mathbf{C}$-worlds (i.e. worlds at which $\mathbf{C}$ is true), neither disjunct is true in *all* of them, meaning both $\mathbf{C} \mathrel{\Box\!\!\rightarrow} \mathbf{I}$ and $\mathbf{C} \mathrel{\Box\!\!\rightarrow} \mathbf{F}$ are false on the Lewisian analysis.

The unified analysis that I defend in this thesis also implies that both $\mathbf{C} \mathrel{\Box\!\!\rightarrow} \mathbf{I}$ and $\mathbf{C} \mathrel{\Box\!\!\rightarrow} \mathbf{F}$ are false. As it turns out, it even implies that $\mathbf{C} \mathrel{\Box\!\!\rightarrow} (\mathbf{I} \vee \mathbf{F})$ is false. This may strike the reader as unintuitive. Yet it seems to me that if Bizet and Verdi had been compatriots, they *might* have both been German (...or Scottish or Chilean or Indonesian or...). More on this in chapter 6.

Another relevant consequence of Lewis's claim that there may be ties in similarity is that his analysis is incompatible with the following purported principle of conditional logic:

Conditional Excluded Middle (CEM) $(P > Q) \vee (P > \neg Q)$.

[5] Here, I am using '**C**' as a constant to refer to the shared antecedent (in reference to the word 'compatriots'). This symbol is not to be confused with '*C*', which is a variable that represents the consequent of a conditional.

Here, 'P' and 'Q' refer to any propositions whatsoever. When applied to counterfactual conditionals in particular, this principle says that either $P \mathrel{\Box\!\!\rightarrow} Q$ or $P \mathrel{\Box\!\!\rightarrow} \neg Q$ is true, for any values of P and Q. As we have just seen, Lewis thinks that the set of closest **C**-worlds includes both **I**-worlds *and* **F**-worlds. More importantly, it therefore includes both **I**-worlds and ¬**I**-worlds. That means that $\mathbf{C} \mathrel{\Box\!\!\rightarrow} \mathbf{I}$ and $\mathbf{C} \mathrel{\Box\!\!\rightarrow} \neg\mathbf{I}$ are both false on Lewis's analysis. Hence, Lewis's view is incompatible with CEM. Of course, Lewis is aware of this. He writes:

> Given Conditional Excluded Middle, we cannot truly say such things as this:
>
> > *It is not the case that if Bizet and Verdi were compatriots, Bizet would be Italian; and it is not the case that if Bizet and Verdi were compatriots, Bizet would not be Italian; nevertheless, if Bizet and Verdi were compatriots, Bizet either would or would not be Italian...*
>
> I want to say this, and think it probably true; my own theory was designed to make it true. But offhand, I must admit, it does sound like a contradiction. (D. K. Lewis, 1973, p. 80)

It follows from the rejection of CEM that the contrary of a conditional is not equivalent to the contradictory. This requires some explanation. Consider the following:

$\mathbf{C} \mathrel{\Box\!\!\rightarrow} \neg\mathbf{I}$ *If Bizet and Verdi had been compatriots, Bizet would* not *have been Italian*.

$\neg(\mathbf{C} \mathrel{\Box\!\!\rightarrow} \mathbf{I})$ *It is* not *the case that if Bizet and Verdi had been compatriots, Bizet would have been Italian*.

The first of these is known as the 'contrary' of $\mathbf{C} \mathrel{\Box\!\!\rightarrow} \mathbf{I}$, whereas the second is known as the 'contradictory'. One might think that the contrary and the contradictory say the same thing, but on Lewis's view, the latter does not imply the former. As we have just seen, Lewis thinks that the set of closest **C**-worlds includes both **I**-worlds and ¬**I**-worlds. That means that $\mathbf{C} \mathrel{\Box\!\!\rightarrow} \mathbf{I}$ and $\mathbf{C} \mathrel{\Box\!\!\rightarrow} \neg\mathbf{I}$ are both false on Lewis's analysis. Hence, $\mathbf{C} \mathrel{\Box\!\!\rightarrow} \neg\mathbf{I}$ (the contrary) is false even though $\neg(\mathbf{C} \mathrel{\Box\!\!\rightarrow} \mathbf{I})$ (the contradictory) is true.

The unified analysis of natural language conditionals that I defend also implies that the contrary of a conditional is not equivalent to the contradictory. The intuition that they *are* equivalent is what is known as a 'scope fallacy'. (More on this in chapter 6.) Like Lewis's analysis, my analysis is incompatible with CEM. This is a consequence of taking into account multiple possible worlds for the evaluation of a single conditional. But whereas Lewis's analysis takes into account all worlds that tie for the title of 'most similar A-world', strict analyses like mine typically take into account many more worlds than this. In the next section, I look at some examples of strict analyses from early on in the literature.

2.3 Strict Analyses

While Whitehead and Russell advocated an interpretation of indicative conditionals in terms of material implication, a common view among their contemporaries was that indicative conditionals express *strict* implication. This is the view that $A \to C$ is true iff 'A and not-C' is impossible, or equivalently, iff 'not-A or C' is necessary. To help us make sense of this view, let us introduce some modal operators: '$\Diamond$' and '$\Box$'. We can interpret these as quantifiers over possible worlds: '$\Diamond P$' means that P is true in at least one world within a given domain, and '$\Box P$' means that P is true in *all* worlds within a given domain. Intuitively, '$\Diamond P$' means that P is *possible*, and '$\Box P$' means that P is *necessary*. These expressions can also be defined in terms of each other:

$$\Box P =^{\mathbf{df}} \neg\Diamond\neg P,$$

$$\Diamond P =^{\mathbf{df}} \neg\Box\neg P.$$

Using these modal operators, we can express strict implication as follows: A strictly implies C iff $\neg\Diamond(A \wedge \neg C)$. Equivalently: A strictly implies C iff $\Box(\neg A \vee C)$. Given the material equivalence thesis stated in section 2.1, we can see that these are both equivalent to:

strict implication A strictly implies C iff $\Box(A \supset C)$.

Since '$\Box$' ranges over different domains in different contexts, we can express the same thing slightly more formally as follows: A strictly implies C relative to a set of worlds, W, iff $A \supset C$ is true in every member of W.

There were several early endorsements of the view that indicative conditionals express strict implication. One of the earliest comes from MacColl, who writes that the negation of 'If he persists in his extravagance he will be ruined' is equivalent to 'He may persist in his extravagance without necessarily being ruined.' (1880, p. 54) Thus, according to MacColl, the negation of the conditional affirms the *possibility* of persisting without being ruined: $\Diamond(A \wedge \neg C)$. That means that the conditional itself must affirm the *im*possibility of persisting without being ruined: $\neg\Diamond(A \wedge \neg C)$. Slightly later, the great American pragmatist C. S. Peirce echoed the same view, writing that an indicative conditional $A \to C$ is true iff '[i]n *any* possible state of things... either [A] is not true, or [C] is true.' (1896, p. 33) A little later still, the view was echoed also by C. I. Lewis in his (1912) and (1914) papers, the latter of which is titled 'The Calculus of Strict Implication'.

While C. I. Lewis in particular has done a great deal to develop the idea that indicative conditionals express strict implication, none of these early writers did very much to *defend* the idea *beyond* developing it. Back then, there was much less need to do so, because the material interpretation was only just gaining traction, and the view that indicative conditionals express strict implication was apparently commonplace. What we do find is the occasional argument against the material interpretation. Consider e.g. the following passage from C. I. Lewis, who argues that the system of logic involving material implication provides

> not rules for drawing inferences at all, but only propositions about the nature of any world to which this system of material implication would apply. *In such a world, the all-possible must be the real, the true must be necessary, the contingent cannot exist, the false must be absurd and impossible, and the contrary to fact supposition must be quite meaningless*. (C. I. Lewis, 1914, p. 244)

Of course, at a time when the material interpretation was just gaining traction and interpretations in terms of strict implication were apparently commonplace, arguments against the former served more or less as arguments for the latter. But nowadays, the landscape has changed: the material interpretation is commonplace, and there are many additional views to contend with. This makes room for a fresh argument in favour of a strict analysis.

Moreover, with the philosophical advancements of the 20^{th} century, one can now be much clearer about *what* is being argued for. Not only have philosophers developed a possible worlds calculus for doing modal logic, a great deal of philosophy has been written with a view to distinguishing different types of *modality*, such as the epistemic modality, the metaphysical modality, and the nomic modality. (More on this in chapter 3.) Different types of strict implication correspond to different types of modality. One way to think of it is as follows: the type of strict implication attributed to a conditional depends on the defining characteristics of W, the set of worlds to which the strict implication is relative. W may be the set of epistemically possible worlds relative to a particular epistemic agent, or the set of metaphysically possible worlds, or the set of nomically possible worlds... Indeed, W may even be the impoverished set of worlds whose only member is the *actual* world, giving us a view with all the same consequences as the material interpretation. (This helps us to make sense of the C. I. Lewis quote above.) The strict analysis that I argue for in chapter 3 is in terms of the metaphysical modality in particular. In other words, I argue that $A \to C$ is true iff A strictly implies C relative to the set of metaphysically possible worlds. One of the ways in which I argue for this strict analysis is by comparing it against analyses in terms of other types of strict implication. This kind of argument was simply much harder to express a hundred years ago, because much of the relevant literature did not yet exist.

One argument that C. I. Lewis (1914, p. 242) does make in favour of his strict analysis is that, despite appearances, it can explain the apparent equivalence between some instances of 'If *A*, then *C*' and 'Either not-*A* or *C*'. Earlier, it was said that (I) seems to be equivalent to 'Either Tolstoy wrote Anna Karenina or someone else did'. As we saw, the material interpretation coheres with this intuition - at least, it does if we assume the standard *extensional* interpretation of a disjunction, such that it is true iff either of its disjuncts is true (even if only contingently[6] so). On the other hand, if we interpret the relevant disjunction as *intensional*, such that it is true iff *necessarily* at least one of its disjuncts is true, then it is just another way of expressing strict implication: $\Box(\neg A \vee C)$. Importantly, there is reason to think that we do sometimes interpret disjunctions as intensional. Suppose, for example, that someone were to say 'You'll either win the game or you'll have a terrible time playing it.' Suppose you then win the game and have a good time playing it, but you believe that you could have had a good time playing it even if you had lost. On this basis, you might reasonably disagree with the initial disjunctive statement. Doing so would suggest an intensional interpretation: even though one of the disjuncts turns out to be true, your grievance is that it *might* have been that *neither* were true. This is not merely to object to the assertion of the disjunction - it is to object to the disjunction itself.

Still, when a disjunction is asserted, it is not always taken to imply that the falsehood of both its disjuncts is impossible. Similarly, when an indicative conditional is asserted, it is not always taken to imply that the conjunction of its antecedent and its consequent's negation is impossible. Suppose I were to say 'If it rains tomorrow, I'll take my umbrella to work', and suppose it does rain, and I do take my umbrella to work. In such a case, I might reasonably take myself to have said something true even though, intuitively, the antecedent did not *necessitate* the consequent. Strict implication is a very high standard to expect all indicative conditionals to meet: if this is what is required to make an indicative conditional true, then most of the indicative conditionals that we assert are false. For reasons that will become clearer as we go on, I think this is a price worth paying. I do concede, however, that it is a price.

Before moving on to the next section, it is worth mentioning one more thing. As well as talking about strict implication, philosophers often use the term 'strict conditional' (see e.g. Lewis (1973, §1.2)). In particular, they tend to use this term to refer to a *type* of conditional. This conveys not just that every token of that type expresses strict implication, but also that every token expresses the *same form* of strict implication. Here is a more formal definition:

strict conditional A *type* of conditional is strict iff: every token of that

[6] A contingently true proposition is one that is true, but not as a matter of necessity. More generally, a contingent proposition is one that is neither necessarily true nor necessarily false.

type is true iff its antecedent strictly implies its consequent relative to a set of worlds, W, where W is the same for all tokens of that type.

According to the analysis that I defend, the natural language conditional is strict.

As well as talking about strict conditionals, philosophers often use the term 'variably strict conditional' (see e.g. Lewis (1973, §1.3)). In the next section, I explain what is meant by this term, and I look at Stalnaker's (1968) theory according to which the natural language conditional is variably strict.

2.4 Variably Strict Analyses

When describing a type of conditional, the term 'variably strict' conveys the claim that every token of that type expresses *a* form of strict implication, but not necessarily the *same* form. Here is a more formal definition:

variably strict conditional A type of conditional is variably strict iff: every token of that type is true iff its antecedent strictly implies its consequent relative to a set of worlds, W, where W *varies* between tokens of that type.

Describing a type of conditional as 'variably strict' is meaningful only if W varies in a way that is neither arbitrary nor ad hoc. One way in which W may systematically vary is by being the set of closest A-worlds, since 'A' is a variable that represents different propositions in different contexts. This points to a different way of expressing the Lewisian analysis of counterfactual conditionals. As we saw above, according to Lewis, $A \mathrel{\Box\!\!\rightarrow} C$ is true iff all the closest A-worlds are worlds at which C is also true. Equivalently: $A \mathrel{\Box\!\!\rightarrow} C$ is true iff all the closest A-worlds are worlds at which $A \supset C$ is true (since $A \supset C$ always has the same truth value as C within the set of A-worlds). Equivalently again: $A \mathrel{\Box\!\!\rightarrow} C$ is true iff A strictly implies C relative to a set of worlds, W, where W varies to include all and only the closest A-worlds. Hence, according to Lewis, the counterfactual conditional is variably strict.

Stalnaker (1968) gives us a slightly earlier theory of conditionals in terms of the same concept, but written before the term 'variably strict' had been coined. Unlike Lewis, Stalnaker aims to capture our usage of both counterfactual conditionals *and* indicative conditionals. He writes:

> Consider a possible world in which A is true, and which otherwise differs minimally from the actual world. *"If A, then* [C]*" is true (false) just in case* [C] *is true (false) in that possible world.* (Stalnaker, 1968, p. 102)

More succinctly: $A > C$ is true iff the closest A-world is a C-world. Since an A-world is a C-world iff it is a world at which $A \supset C$ is true, we can rephrase Stalnaker's view (less succinctly) as follows: $A > C$ is true iff A strictly implies C relative to a set of worlds, W, where W varies to include only the closest A-world. Hence, if Stalnaker is right, the natural language conditional is variably strict.

It ought to be clear that Stalnaker's account has much in common with Lewis's. However, there are also some key differences. One of the most important of these is that Stalnaker does not allow for ties in closeness. Rather, on Stalnaker's view, possible worlds are well-ordered: as long as there are *some* A-worlds, there will always be one that is uniquely close to the actual world. This is called the *uniqueness assumption*, and it allows/forces Stalnaker to accept CEM. Accepting CEM is appealing on the face of it, but as we saw in section 2.2, there are some pairs of contrary conditionals that do not seem to conform to CEM. In particular, Stalnaker's view seems to imply that exactly one of the Bizet/Verdi conditionals is true. Stalnaker (1981) has a response to this, which is to invoke the rather technical notion of *supervaluations* à la Van Fraassen (1966). However, it is not at all clear to me that the result is more appealing than simply rejecting the principle, as I will explain.

Supervaluations are assignments of truth values or truth value gaps to vague sentences. They are determined by the truth values of all the propositions expressed by all the ways of arbitrarily eliminating the vagueness in question. As Stalnaker explains:

> A sentence is *true* according to a supervaluation if and only if it is true on *all* corresponding classical valuations, *false* if and only if it is false on *all* corresponding classical valuations and neither true nor false it [sic] it is true on some of the classical valuations and false on others. (Stalnaker, 1981, p. 90)

This effectively means that Stalnaker can accept CEM while claiming that the Bizet/Verdi conditionals are neither true nor false. He writes:

> On Lewis's [analysis]... both counterfactuals are false. On the analysis I am defending, both are indeterminate - neither true nor false. It seems to me that the latter conclusion is clearly the more natural one. I think most speakers would be as hesitant to deny as to affirm either of the conditionals, and it seems as clear that one cannot deny them both as it is that one cannot affirm them both. (Stalnaker, 1981, p. 92)

This may *sound* like having one's cake and eating it, but to be clear, what Stalnaker is saying is that $(P > Q) \vee (P > \neg Q)$ is a valid logical principle even though, for some values of P and Q, neither $P > Q$ nor $P > \neg Q$ is true. Ultimately then, however we handle the Bizet/Verdi conditionals, we are going to be committed to some unintuitive claims.

As mentioned above, Stalnaker's analysis is intended to apply to both indicative and counterfactual conditionals, making it a *unified* theory of conditionals. In other words, according to Stalnaker, both indicative and counterfactual conditionals ultimately express the same kind of proposition. The theory that I defend is the same in this regard. For that reason, in the next section, I explore this idea in more detail.

2.5 Unified Theories

Consider again the following conditionals:

(I) *If Tolstoy did not write Anna Karenina, then someone else did.*

(II) *If Tolstoy* had *not written Anna Karenina, then someone else* would *have.*

Pairs of conditionals like (I) and (II) are often referred to as 'corresponding' indicative and counterfactual conditionals, the idea being that their constituent propositions seem to be the same even though the words used to express them are different:

correspondence 'If P, then Q' corresponds to 'If R, then S' iff $P = R$ and $Q = S$.

Of course, it is by no means *certain* that the constituent propositions of (I) and (II) are the same, but there is at least some intuitive pull towards the idea, since they seem to be about the same states of affairs. Intuitively, however, (I) and (II) are logically distinct. At the very least, we react very differently to them. This is a key motivation for the idea that indicative and counterfactual conditionals involve different logical connectives: if they have the same constituent propositions but are nonetheless logically distinct, then they must have different logical connectives. (See e.g. Lewis (1973, p. 3). More on this in chapter 5.)

As well as explaining our different reactions to conditionals like (I) and (II), a theory of conditionals should identify what these propositions have in common with each other. In other words, a theory of conditionals should account for both the difference *and* the similarity between indicative and counterfactual conditionals. Cf. Bennett:

> We want a good analytic understanding of each of the two kinds of conditional. One might hope to find a Y-shaped analysis of them - first stating what is common to the two kinds and then bifurcating in order to describe the differences. That they have much in common seems clear. (Bennett, 2003, p. 8)

Lewis's overall view of conditionals gives us an example of a Y-shaped analysis. As we saw earlier, Lewis's analysis of counterfactual conditionals implies that they behave like material conditionals whenever their antecedents are true; what we did not see earlier is that Lewis thinks *indicative* conditionals behave like material conditionals whether their antecedents are true or false: that is, he endorses the material interpretation of indicative conditionals (see Lewis (1973, p. 72, fn)). Hence, on Lewis's view, the truth conditions of indicative and counterfactual conditionals are the same whenever their antecedents are true, and different whenever their antecedents are false. This places the point of divergence (i.e. the "fork" of the Y) within the truth conditions themselves.

Unified theories place the point of divergence elsewhere. In other words, they attribute the same truth conditions to both indicative and counterfactual conditionals (if they attribute truth conditions at all). That does not mean, however, that they cannot explain our different reactions to (I) and (II). Consider again Stalnaker's unified theory of conditionals. According to Stalnaker (1968, p. 98), indicative and counterfactual conditionals have the same truth conditions in virtue of expressing the same kind of proposition: a function which picks out the most similar A-world. However, according to Stalnaker, similarity depends on *context*, meaning the conditional function picks out different A-worlds even when A is held constant.

The above gives Stalnaker a way of denying that corresponding indicative and counterfactual conditionals are logically equivalent while affirming that they have the same truth conditions. In this respect, the unified analysis that I defend is different: it implies that corresponding indicative and counterfactual conditionals are logically equivalent. As unintuitive as this idea may at first seem, there is some support for it in the literature. For example, Ellis defends a unified theory of conditionals according to which (I) and (II)

> differ from each other in tense and mood, and hence in the circumstances in which it would be appropriate to assert them. But they contain the same antecedent supposition and the same consequent, and the same considerations are relevant to the acceptability or otherwise of the belief they both express. (Ellis, 1978, p. 120)

Ellis's claim is that conditionals like (I) and (II) have different *assertability conditions* in virtue of the different words used to express them. In chapter 4, I argue for a similar claim: that the words used to express these conditionals suggest different sets of common knowledge, and these different sets of common knowledge have different impacts on assertability. This can explain our different reactions to conditionals like (I) and (II).

Unified theories may clash with some intuitions, but they also cohere with others. For example, there is intuitively some sort of connection between a conditional $A > C$

and the conditional probability of C given A. With this in mind, Stalnaker (1970b) argues that the probability of $A > C$ is equal to the conditional probability of C given A (except where the conditional probability is undefined). This thesis is known as 'Stalnaker's Thesis', and we can represent it in symbols as follows:

Stalnaker's Thesis $\mathrm{P}(A > C) = \mathrm{P}(C|A)$, except where $\mathrm{P}(A) = 0$.

Here, '$\mathrm{P}(A > C)$' represents the probability of $A > C$, and '$\mathrm{P}(C|A)$' represents the conditional probability of C given A. The rationale behind the qualification is that $\mathrm{P}(C|A)$ is undefined when $\mathrm{P}(A) = 0$.[7]

Stalnaker's Thesis is elegantly simple, so it would be nice if it were true. Unfortunately, Lewis (1976) argues so persuasively against it that even Stalnaker has now abandoned it. In particular, Lewis shows that no proposition can have a probability equal to this conditional probability unless the language in which it is expressed is a *trivial* one, i.e. one that is severely limited in terms of what other propositions it can express. (More on this in chapter 7.) Still, it remains intuitively plausible that there is *some* sort of connection between $A > C$ and $\mathrm{P}(C|A)$. To borrow a phrase from Gilbert Ryle (1949, pp. 110-11), conditionals are 'inference-tickets' from premises to conclusions: $A > C$ licenses the inference from A to C (by *modus ponens*), as well as the inference from $\neg C$ to $\neg A$ (by *modus tollens*). The most obvious way to evaluate $A > C$ qua inference-ticket is in terms of the conditional probability of C given A, which is the same no matter what '>' represents. Hence, any theory according to which some values of A and C yield a true indicative conditional but a false counterfactual (or vice versa) must explain why this difference in truth value does not imply a difference in value qua inference-ticket. On the other hand, no such explanation is necessary on a theory of conditionals according to which corresponding indicative and counterfactual conditionals have the same truth values. For this reason, unified theories are particularly well-placed to make sense of the intuitive connection between conditionals and conditional probability.

It is very surprising that Stalnaker's Thesis is false. If a conditional does not express a proposition whose probability is equal to the conditional probability of its consequent given its antecedent, then what kind of proposition *does* it express? Intuitively, a conditional's truth conditions should be related in *some* way to the relevant conditional probability. With this in mind, I argue (in chapter 7) that we can use my analysis to make sense of the idea that $A > C$ is true iff the objective conditional probability of C given A is 1. Not only does this idea do justice to the intuition that there is some sort of connection between conditionals and conditional probability, it does so in a way that is consistent with the idea that conditionals express propositions: as I show in chapter 7, Lewis's argument against Stalnaker's Thesis does not threaten the connection that I propose.

[7] As Hájek (2003) points out, this rationale can be challenged. More on this in chapter 7.

Some theorists have given up on the hunt for a conditional proposition, in part because it seems like the probability of the conditional *should* be equal to the relevant conditional probability. In other words, if conditionals do not express a proposition whose probability is equal to the relevant conditional probability, then maybe they do not express a proposition at all. This will be the subject of the next section. To finish off this section, however, let us consider one more argument in favour of unified theories.

In his (1947), Goodman argues that counterfactual conditionals can be paraphrased as indicative conditionals by performing a logical operation known as 'contraposition'. To contrapose a conditional is to swap and negate its antecedent and consequent, so that $P > Q$ becomes $\neg Q > \neg P$. It is a widely accepted principle of conditional logic that this operation preserves truth value.[8] In symbols:

Contraposition $P > Q \equiv \neg Q > \neg P$.

According to Goodman, 'any counterfactual can be transposed into a conditional with a true antecedent and consequent.' (1947, p. 114) For example, on Goodman's view, contraposing (II) gives us:

(II)* Since *no one else wrote Anna Karenina, Tolstoy did*.

It is of course a somewhat controversial claim that (II)* is the contrapositive of (II), in part because (II)* does not even feature the word 'if'. Still, this is a hill Goodman is willing to die on:

> That 'since' occurs in the contrapositive shows that what is in question is a certain kind of connection between the two component sentences; and the truth of statements of this kind–whether they have the form of counterfactual or factual conditionals or some other form–depends not upon the truth or falsity of the components but upon whether the intended connection obtains. (Goodman, 1947, p. 10)

2.6 NTV Theories

Like me, many philosophers who theorise about conditionals nowadays are motivated by a dissatisfaction with the material interpretation of indicative conditionals. For example, Edgington writes:

> [T]he truth-functional account [has] intolerable consequences, and we have not seen a way to make them tolerable. There is a solution... but it lies ahead. (Edgington, 1995, p. 247)

[8] This is not to say that there are no analyses of conditionals on which the principle fails. Indeed, it is generally thought to fail for counterfactual conditionals.

Similarly, Bennett writes:

> The horseshoe analysis[9] of → should be rejected, because of the failure... to reconcile it with the intuitive data... The horseshoe analysis having failed, we must look further. (Bennett, 2003, pp. 43-45)

However, for many of these theorists, the main cause of dissatisfaction is not that the material interpretation implies the logical impossibility of certain intuitively true propositions; rather, the main cause of dissatisfaction is the material interpretation's incompatibility with an intuitive idea that Gibbard calls the 'Ramsey test thesis'.

In a surprisingly famous footnote, Ramsey (1931a, p. 247, footnote 1) identifies what he deems to be the deliberative procedure by which one comes to accept or reject a conditional. That deliberative procedure is simply a matter of fixing one's degree of belief in the consequent given the antecedent. In other words, according to Ramsey, it is a matter of determining the relevant *subjective conditional probability*. This is a highly intuitive and widespread idea - Ramsey may have written it down, but he presumably did not *originate* it. Nonetheless, this deliberative procedure has come to be known as the 'Ramsey test', and the general idea behind it is what Gibbard calls the 'Ramsey test thesis'. He writes:

> [B]y the *Ramsey test thesis*, I mean the thesis that, in whatever ways the acceptability, assertability, and the like of a proposition depend on its subjective probability, the acceptability, assertability, and the like of an indicative conditional... depend on the corresponding subjective conditional probability... (Gibbard, 1980a, p. 253)

The material interpretation of indicative conditionals seems to be incompatible with the Ramsey test thesis: if indicative conditionals are material, then their acceptability seems to depend not on the subjective conditional probability of the consequent given the antecedent, but on whether the antecedent is false or the consequent is true. This is bad news for the material interpretation, and with the inadequacy of its truth-functional truth conditions in mind, Ernest Adams (1965, 1975) developed a system of logic according to which indicative conditionals have 'conditions of 'justified assertability' rather than conditions of truth'. (1965, p. 172) In particular, Adams built this system of logic around the idea that the justified assertability of an indicative conditional is equal to the subjective conditional probability of its consequent given its antecedent. This thesis has come to be known as 'Adams' Thesis', and we can represent it in symbols as follows:

Adams' Thesis $\mathrm{As}(A \to C) = \mathrm{P}(C|A)$, except where $\mathrm{P}(A) = 0$.

[9] Bennett refers to the material interpretation of indicative conditionals as the 'horseshoe analysis' because of the '⊃' symbol used to represent material implication. I prefer the term 'material interpretation' just because there is some doubt as to whether '→ is ⊃' qualifies as an analysis.

Precisely what Adams meant by 'justified assertability' is debatable (see Hájek (2012a)), but what is clear is that he did not mean *truth* conditions. Following this, Gibbard (1980b), Appiah (1985), Edgington (1986), and Bennett (1988, 2003) have all been led in the same direction, endorsing theories according to which indicative conditionals are *non-truth-evaluable*. Following Lycan (2001), I use the initialism 'NTV' (No Truth Value) to refer to these theories.

The Ramsey test thesis has led people to NTV theories partly because of its role in a type of thought experiment known as a 'Gibbardian stand-off'. Gibbard (1980b, p. 231) describes a situation involving two people who have very good grounds (according to the Ramsey test thesis) for asserting contrary conditionals. As we saw above, these are conditionals with the same antecedent but contradictory consequents. Here is how the thought experiment goes. There is a game of poker in which two players ('Sly Pete' and 'Mr Stone') have not yet folded. There are also two onlookers. Onlooker 1 sees Mr. Stone's hand and makes a gesture to Sly Pete to communicate the content of Mr. Stone's hand. Additionally, Onlooker 1 knows that Sly Pete will act on this information rationally, calling a bet only if his own hand is better. Onlooker 2 sees both Sly Pete's hand *and* Mr. Stone's hand, the latter of which is better than the former. Mr. Stone suspects that something fishy is going on, so he orders everyone except Sly Pete to leave the room. Afterwards, without gaining any relevant new information, Onlooker 1 has good grounds (according to the Ramsey test thesis) to assert:

$\mathbf{B} \rightarrow \mathbf{W}$ *If Sly Pete called a bet, then he won.*

After all, the subjective conditional probability of **W** given **B** is high from Onlooker 1's perspective. On the other hand, Onlooker 2 has good grounds (according to the Ramsey test thesis) to assert:

$\mathbf{B} \rightarrow \neg\mathbf{W}$ *If Sly Pete called a bet, then he did* not *win.*

After all, the subjective conditional probability of $\neg$**W** given **B** is high from Onlooker 2's perspective. Since neither onlooker has any relevant false beliefs, and since (according to Gibbard) '[o]ne sincerely asserts something false only when one is mistaken about something germane' (1980b, p. 231), it seems as though both onlookers assert true propositions *if* they assert propositions at all. But the idea that contrary conditionals can both be true contradicts the following purported principle of conditional logic:

Conditional Non-Contradiction (CNC) $\neg((P > Q) \wedge (P > \neg Q))$.[10]

[10] This principle is sometimes restricted to conditionals with antecedents that are *possible* in some sense. However, as we will see in the next chapter, there are many different types of possibility. For simplicity's sake, I will stick to this unrestricted formulation of CNC.

For this reason, Gibbard concludes that indicative conditionals do not express propositions. In other words, he concludes that they are *non-truth-evaluable*.

Bennett agrees with Gibbard, writing:

> [Contrary] conditionals cannot both be true, because their being so would conflict with the principle of Conditional Non-Contradiction [CNC]... which is almost indisputably true. According to the horseshoe analyses, CNC is false... But on no other account of indicative conditionals has CNC any chance of coming out false. (Bennett, 2003, p. 84)

Whether this last claim of Bennett's is true depends on whether CNC is restricted to conditionals with antecedents that are *possible* in some sense. If not, then CNC has every chance of coming out false on a strict analysis of indicative conditionals: if it is impossible that P, then it is also impossible that P and Q, and it is *also* impossible that P and not-Q; hence, on a strict analysis, if $\neg\Diamond P$, then both $\Box(P \supset Q)$ and $\Box(P \supset \neg Q)$ are (vacuously) true.

Bennett adds that contrary conditionals like $\mathbf{B} \to \mathbf{W}$ and $\mathbf{B} \to \neg\mathbf{W}$ cannot both be *false* either, because this 'implies that countless conditionals that would ordinarily be thought to be acceptable are actually false'. (2003, p. 94) To say they are both false is to reject CEM for indicative conditionals. I say they are both false, and I think that many other conditionals that are commonly thought to be acceptable are false as well. I argue (in chapter 4) that the assertion of a false conditional may serve its purpose just as well as the assertion of a true conditional. For this reason, I also reject Gibbard's claim that '[o]ne sincerely asserts something false only when one is mistaken about something germane'. (1980b, p. 231)

The claim that false conditionals can be assertable is bound to ruffle many contemporary philosophers' feathers. Nonetheless, I think it is a price worth paying in order to preserve the idea that indicative conditionals are truth-evaluable. As Rieger writes:

> In arguing for one theory over another, one is not making a descriptive claim about how speakers of natural language use conditional sentences. No such approach can be expected to result in a consistent theory. Rather, one is making a normative proposal, as to the places in which intuition should be respected and those in which it must be given up. (Rieger, 2013, p. 3172)

To argue that indicative conditionals are non-truth-evaluable is to make the normative proposal that we should give up on finding them a suitable set of truth conditions. I think accepting this proposal should be a last resort. If indicative conditionals do not express propositions, then conditional thoughts cannot be shared, because propositions - the bearers of truth values - are necessary to make sense of the idea that two thinkers can think the *same* thought. Of course, I do not mean to suggest that those who endorse

NTV theories do so impulsively - on the contrary, they present impressive arguments in favour of their theories. However, as we have seen, I disagree with many of the premises in those arguments. The following chapters are an attempt to explain *why*.

2.7 Conclusion

In this chapter, I have given an evaluative overview of some of the most influential theories of conditionals. In the course of doing so, I have explained some concepts and arguments from the literature that will be key ingredients in the following chapters. I hope the reader is now in a good position to understand not only the content of those chapters, but the motivation for writing them. I also hope the reader is in a good position to evaluate the theory that I defend. As we have seen, theories of conditionals invariably come at the cost of *some* intuitions. Now that we have a general sense of the marketplace, let us see whether the cost of a strict analysis is a price worth paying.

Chapter 3

A Modal Analysis of Indicative Conditionals

The Ellisian paradox that we saw in section 2.1 shows that the material interpretation of indicative conditionals has alarming consequences. In particular, it shows that the material interpretation makes some intuitively true propositions logically impossible. I take the avoidance of this to be a desideratum of a theory of indicative conditionals, and I think this desideratum can be met without giving up truth conditions. A modal analysis seems to be the obvious solution. As Ryle once wrote:

> An 'if-then' sentence can nearly always be paraphrased by a sentence containing a modal expression, and *vice versa*. Modal and hypothetical sentences have the same force. (Ryle, 1949, p. 111)

In this chapter, I explain and defend the view that indicative conditionals are metaphysically strict, meaning $A \rightarrow C$ is true iff $A \supset C$ is true in all metaphysically possible worlds. I begin by elaborating on the desideratum mentioned above, which (for reasons that will become clear) I call the 'minimal respect desideratum'. I then explain in detail why the material interpretation does not meet this desideratum. In sections 3.3 to 3.5, I consider strict analyses resulting from what I take to be the three forms of strict implication most plausibly expressed by indicative conditionals: epistemically strict implication, metaphysically strict implication, and nomically strict implication (i.e. strict implication relative to worlds that share the same laws of nature as the actual world). All three strict analyses have virtues, but only the metaphysically strict analysis and the nomically strict analysis meet the minimal respect desideratum. In section 3.6, I argue that the nomically strict analysis faces an objection that the metaphysically strict analysis can easily avoid: given a Humean view of the laws of nature, a vicious circularity arises in which the facts of the future help to determine the present history-to-future conditional truths.

None of this constitutes a conclusive argument in favour of the metaphysically strict analysis - I certainly do not expect the reader to embrace the analysis on the basis of this chapter alone. Rather, my more modest aim in this chapter is to show that, in at least some respects, the metaphysically strict analysis outperforms at least some of its biggest competitors. For this reason, I think it deserves our consideration.

3.1 The Minimal Respect Desideratum

Even in philosophy, it is impossible not to have some starting assumptions. One of the things that I take for granted in this thesis is that a theory of conditionals is an attempt to identify universal laws of logic, i.e. laws that are true for everyone throughout the actual world, and true even throughout the many possible worlds of Lewis's modal realism (see section 2.2), if such worlds exist. In this respect, the laws of logic are like the laws of mathematics: they are as general as it gets.

Another thing that I take for granted is that a theory of conditionals must respect our intuitions to some extent, not because we have an ability to intuit the structure of the universe, but because, as Rieger puts it, we are 'making a normative proposal, as to the places in which intuition should be respected and those in which it must be given up'. (2013, p. 3172) Indeed, to my mind, it is precisely this normative element that makes the laws of logic universal.

As the previous chapter demonstrates, one ought not to expect a consistent theory of conditionals to imply the truth (falsehood) of every relevant intuitively true (false) proposition. Yet there are different degrees of respect, and it seems to me that one might reasonably expect a consistent theory of conditionals to pay at least *minimal* respect to our intuitions. That is, one might reasonably expect to find a consistent theory that avoids making any intuitively true propositions *logically impossible*, thereby rendering the relevant intuitions radically defective, inconsistent with even the most general laws. With this in mind, I suggest the following desideratum:

minimal respect desideratum A theory of conditionals should not make a proposition logically impossible if the proposition is intuitively true.

Here, 'proposition' may refer to a conditional proposition *or* an unconditional proposition, since some unconditional propositions are affected by theories of conditionals in virtue of featuring conditionals as component parts or being about conditionals in some way.

Question: Intuitively true *to whom*? Answer: Competent users of the relevant language. Of course, an advocate of a theory of conditionals might argue that one is competent in this regard only if one's intuitions are minimally respected by the theory being advocated, but this would be to delineate the domain in an ad hoc way. I take for granted that most native English speakers are competent users of conditionals, since 'if' is one of our most common words. More importantly, I take for granted that the reader is a competent user of the relevant language - it is, after all, the reader's intuitions to which I am appealing.

Having clarified this, it will be worth taking a moment also to clarify the concept of logical impossibility. It is not at all obvious that such a fundamental concept can be analysed in fully independent terms, but what we can say is that *P* is logically impossible iff the laws of logic suffice to make it false.[1] Since a theory of conditionals is an attempt to identify some of the laws of logic, its truth has consequences regarding logical impossibility. More specifically, we can say that a theory of conditionals makes *P* logically impossible iff its conjunction with the other laws of logic suffices to make *P* false.

Some paradigm examples of the laws of logic include those that govern negations, conjunctions, and disjunctions. While I reject the classical (i.e. material) interpretation of 'if', I nonetheless assume the classical, truth-functional interpretations of 'not', 'and', and 'or'.[2] Classical propositional logic models the behaviour of these logical constants[3] by specifying rules regarding the assignment of truth values to formulae. These are the *formal semantics* of the system, and they represent (at least some of) the laws of logic. The rules for '¬', '∧', and '∨' are as follows:

P	Q	$\neg P$	$P \wedge Q$	$P \vee Q$
T	T	F	T	T
T	F	F	F	T
F	T	T	F	T
F	F	T	F	F

[1] Similarly *P* is logically *possible* iff the laws of logic do *not* suffice to make it false, and *P* is logically *necessary* iff the laws of logic suffice to make it *true*.

[2] As per section 2.3, I actually think an *intensional* interpretation of 'or' may sometimes be preferable, but that is a logical battle for another day.

[3] As Sider notes, it is not immediately obvious how to distinguish the category of logical constants:

> '[L]ogicians do not focus on just any old phrases. They focus on 'and', 'or', 'not', 'if. . . then', and so on... [These are] the words for which they introduce special symbolic correlates... [They are] the *logical constants*... [T]he fact is that logicians do not treat 'bachelor' and 'unmarried' as logical constants... But... why don't they? What's so special about 'and', 'or', 'all', and 'some'?... Why not expand logic... to include the logic of bachelorhood and unmarriage?... [T]here's no formal obstacle to doing just that.' (Sider, 2010, p. 10)

Nonetheless, 'not', 'and', 'or', and 'if' are all paradigmatic examples of logical constants.

For an instructive example of a proposition that the laws of logic suffice to make false, note that there is no assignment of truth values according to which $P \wedge Q$ is true and $P \vee Q$ is false; hence, on the assumption that the classical interpretations of conjunctions and disjunctions are correct, the laws of logic suffice to make false any proposition of the form '*P* and *Q* and it is not the case that *P* or *Q*', or less awkwardly:

truth value claim 1 *It is true that P and Q, and it is false that P or Q.*

I take for granted that a proposition can have its own truth value while being about the truth values of other propositions. If the reader disagrees, simply delete any instances of 'it is true that' and replace any instances of 'it is false that' with 'it is not the case that'. The result will be a more awkward claim, but not to the point of unintelligibility.

I trust that the above helps to justify and illuminate the minimal respect desideratum. In the next section, I demonstrate that the material interpretation of indicative conditionals fails to meet this desideratum by revisiting the Ellisian paradox first introduced in section 2.1.

3.2 The Ellisian Paradox Revisited

Consider again the thought experiment from section 2.1, in which the Portuguese and Qatari national football teams are competing against each other in a game where they must keep playing until a winner is established. The game comes to an end. Now consider:

$\neg\mathbf{P} \rightarrow \neg\mathbf{Q}$ *If Portugal did not win the game, then Qatar did not win it (either).*

$\mathbf{P} \rightarrow \mathbf{Q}$ *If Portugal won the game, then Qatar won it (too).*

Those of us who understand the setup of the thought experiment are likely to have the intuition that the following claim is true:

truth value claim 2 *It is false that if Portugal did not win the game, then Qatar did not win it (either), and it is false that if Portugal won the game, then Qatar won it (too).*

But if the material interpretation is true, then it cannot be the case that $\neg\mathbf{P} \rightarrow \neg\mathbf{Q}$ and $\mathbf{P} \rightarrow \mathbf{Q}$ are both false, because there is no assignment of truth values according to which $\neg\mathbf{P} \supset \neg\mathbf{Q}$ and $\mathbf{P} \supset \mathbf{Q}$ are both false, as demonstrated by the following:

$\mathbf{P}$	$\mathbf{Q}$	$\mathbf{P \supset Q}$	$\mathbf{\neg P}$	$\mathbf{\neg Q}$	$\mathbf{\neg P \supset \neg Q}$
T	T	T	F	F	T
T	F	F	F	T	T
F	T	T	T	F	F
F	F	T	T	T	T

The conjunction of the material interpretation and the other laws of logic therefore suffices to make truth value claim 2 false. Hence, the material interpretation makes truth value claim 2 logically impossible.

Now consider again the following set of conditionals, inspired by Ellis (1978, p. 118):

$\mathbf{G} \rightarrow \mathbf{S}$ *If I was born in Glasgow, then I was born in Scotland.*

$\mathbf{G} \rightarrow \mathbf{E}$ *If I was born in Glasgow, then I was born in England.*

$\mathbf{L} \rightarrow \mathbf{S}$ *If I was born in London, then I was born in Scotland.*

Those of us who know the locations of Glasgow and London are likely to have the intuition that the following claim is true:

truth value claim 3 *It is true that if I was born in Glasgow, then I was born in Scotland, and it is false that if I was born in Glasgow, then I was born in England, and it is false that if I was born in London, then I was born in Scotland.*

But if the material interpretation of indicative conditionals is true, then it cannot be the case that $\mathbf{G} \rightarrow \mathbf{S}$ is true and $\mathbf{G} \rightarrow \mathbf{E}$ and $\mathbf{L} \rightarrow \mathbf{S}$ are both false, as demonstrated by the following:

$\mathbf{G}$	$\mathbf{S}$	$\mathbf{E}$	$\mathbf{L}$	$\mathbf{G \supset S}$	$\mathbf{G \supset E}$	$\mathbf{L \supset S}$
T	T	T	T	T	T	T
T	T	T	F	T	T	T
T	T	F	T	T	F	T
T	T	F	F	T	F	T
T	F	T	T	F	T	F
T	F	T	F	F	T	T
T	F	F	T	F	F	F
T	F	F	F	F	F	T
F	T	T	T	T	T	T
F	T	T	F	T	T	T
F	T	F	T	T	T	T
F	T	F	F	T	T	T
F	F	T	T	T	T	F
F	F	T	F	T	T	T
F	F	F	T	T	T	F
F	F	F	F	T	T	T

As can be seen, there is no assignment of truth values according to which **G** $\supset$ **S** is true and **G** $\supset$ **E** and **L** $\supset$ **S** are both false. Hence, the material interpretation makes truth value claim 3 logically impossible as well.

This should suffice to establish that the material interpretation of indicative conditionals does not meet the minimal respect desideratum. In the next section, I consider a strict analysis of indicative conditionals - in particular, one according to which indicative conditionals express a form of strict implication relative to a set of *epistemically* possible worlds. I argue that this analysis fails to meet the minimal respect desideratum because, like the material interpretation, it makes truth value claim 3 logically impossible.

3.3 The Epistemically Strict Analysis

Indicative conditionals often seem to have an epistemic flavour. As we saw in section 2.6, it is a popular idea that 'the acceptability, assertability, and the like of an indicative conditional... depend on the corresponding subjective conditional probability...' (Gibbard, 1980a, p. 253) This idea, which Gibbard calls the 'Ramsey test thesis', is captured by the Gibbardian stand-off: a situation in which contrary conditionals are asserted by two different people, but there does not seem to be anything wrong with either assertion given the information that each person has. Thought experiments like this make it tempting to think that indicative conditionals are subjective. That is, they make it tempting to think that $A \rightarrow C$ is true/false *for a subject*, S, depending on S's situation. One way to capture this idea is to suppose that indicative conditionals express a form of strict implication relative to the set of worlds that are epistemically possible from S's point of view. In other words: $A \rightarrow C$ is true for S iff $A \supset C$ is epistemically necessary for S.

To make proper sense of this, we need to have a clear grasp of the epistemic modality. Whilst P is logically impossible iff the laws of logic suffice to make it false, it is not the case that P is epistemically impossible iff "the laws of epistemology" suffice to make it false - in general, the field of epistemology does not purport to specify laws about what can or cannot be true. Rather, the modality in question is simply one that relates to a subject's epistemic situation. For example, Stalnaker defines epistemic possibilities as those that are 'consistent with the subject's knowledge.' (1970b, p. 68) One might want to swap 'knowledge' for 'beliefs' (or 'justified beliefs', or 'true beliefs'...), but the general idea of consistency with the subject's epistemic situation should remain the same.

Since there are different modalities, the word 'consistent' is ambiguous. Normally, when a philosopher talks about consistency without specifying the type of modality in question, a reasonable assumption is that they mean logical consistency, since the concept of consistency is central to the study of logic. With this in mind, let us say that P

is epistemically possible for S iff it is *logically* consistent with S's knowledge set, X. From this, it follows that P is epistemically *im*possible for S iff it is logically *in*consistent with X. And it also follows that P is epistemically *necessary* for S iff $\neg P$ is logically inconsistent with X, i.e. iff X logically implies (or *entails*) P.[4] With this in mind, we can express the epistemically strict analysis as follows:

epistemically strict analysis $A \rightarrow C$ is true for S iff X logically implies $A \supset C$, where X is S's knowledge set.

For many, the fact that this analysis makes the truth of a conditional relative to a subject will already be enough of a reason to dismiss it. However, some of the early proponents of strict analyses seem to have had something like the epistemically strict analysis in mind. Consider, for example, the following passage from C. S. Peirce:

> '[T]he Philonian logicians have always insisted upon beginning the study of conditional propositions by considering what such a proposition means in a state of omniscience... Duns Scotus terms such a conditional proposition a "*consequentia simplex de inesse*"... The consequence *de inesse*... is expressed by... saying... "Either [A] is not true or [C] is true." But an *ordinary* Philonian conditional is expressed by saying, "In *any* possible state of things... either [A] is not true, or [C] is true."' (Peirce, 1896, p. 33)

Peirce is rejecting the identification of the natural language conditional with the material conditional, arguing that they are equivalent only when used by someone who is omniscient. This claim makes sense if we analyse conditionals in terms of an epistemic form of strict implication. First, note that the epistemically strict analysis is actually *variably* strict (see section 2.4): on the epistemically strict analysis, if $A \rightarrow C$ is true, then A expresses strict implication relative to a set of worlds, W, that *varies* depending on the subject's knowledge set, X. Second, note that every false proposition is logically inconsistent with a true proposition (its negation), so for an omniscient S, W is the set of worlds whose only member is the actual world. Lastly, note that if W is the set of worlds whose only member is the actual world, then a strict analysis simply reduces to material implication. So if the epistemically strict analysis is true, then $A \rightarrow C$ is equivalent to $A \supset C$ for an omniscient S. For those of us whose knowledge is limited, however, $A \rightarrow C$ is *not* equivalent to $A \supset C$, because W includes many other possible worlds besides the actual world.

[4] There are interesting variations of this definition in the literature. For example, Kment writes that 'P is *epistemically necessary for an agent A* just in case the empirical evidence A possesses and ideal reasoning (i.e., reasoning unrestricted by cognitive limitations) are sufficient to rule out ~P.' (2021, §1) Depending on how one conceives of empirical evidence and ideal reasoning, this may turn out to be equivalent to our definition, but even if not, it should still suffice to make the argument made in this section.

This quote of Peirce's points to two virtues of a strict analysis in terms of epistemic possibility. One virtue is that the analysis helps to explain what is appealing about the material interpretation of indicative conditionals: the material interpretation gets things right *if* one is omniscient. The other virtue is that the analysis helps to explain what is unsatisfactory about the material interpretation: the material interpretation gets things right *only if* one is omniscient.

Despite these virtues, I think we should reject the epistemically strict analysis, because it does not meet the minimal respect desideratum. If the epistemically strict analysis of conditionals is true, then $\mathbf{G} \rightarrow \mathbf{S}$ is true for S iff X logically implies $\mathbf{G} \supset \mathbf{S}$, where X is S's knowledge set. But if X logically implies $\mathbf{G} \supset \mathbf{S}$, then it also logically implies $(\mathbf{G} \supset \mathbf{E}) \vee (\mathbf{L} \supset \mathbf{S})$, since there is no assignment of truth values that makes the first conditional true without making at least one of the other conditionals true as well. In other words, it is logically impossible that $\mathbf{G} \rightarrow \mathbf{S}$ is true (for S) and $\mathbf{G} \rightarrow \mathbf{E}$ and $\mathbf{L} \rightarrow \mathbf{S}$ are both false (for S). The epistemically strict analysis therefore makes truth value claim 3 logically impossible.

What if we had started with a different definition of epistemic possibility? Suppose, for example, that we were to say that P is an epistemic impossibility for S iff P is *metaphysically* inconsistent with X. Still, the analysis will fail to meet the minimal respect desideratum. If we incorporate this into a strict analysis of conditionals, then $\mathbf{G} \rightarrow \mathbf{S}$ will be true for S iff X *metaphysically* implies $\mathbf{G} \supset \mathbf{S}$. But if X metaphysically implies $\mathbf{G} \supset \mathbf{S}$, then it also metaphysically implies $(\mathbf{G} \supset \mathbf{E}) \vee (\mathbf{L} \supset \mathbf{S})$, since there is no assignment of truth values that makes the first conditional true without making at least one of the other conditionals true as well. So the same result follows: truth value claim 3 is logically impossible.

In fact, we get the same result no matter what type of inconsistency we appeal to in the definition of epistemic possibility: anything that ϕ-implies $\mathbf{G} \supset \mathbf{S}$ also ϕ-implies $(\mathbf{G} \supset \mathbf{E}) \vee (\mathbf{L} \supset \mathbf{S})$, whatever the value of ϕ. Nor will it help to propose a definition of epistemic impossibility in terms of S's belief set (or justified belief set, or true belief set...), because the argument does not hinge in any way on the propositions in X being *known*. Whichever way we spin it, the view that conditionals express an epistemic form of strict implication does not meet the minimal respect desideratum.

3.4 The Metaphysically Strict Analysis

The main purpose of this section is to introduce the theory that I defend throughout this thesis: namely, the metaphysically strict analysis of indicative conditionals. Before doing that, however, it will be helpful to give the reader a general sense of what the metaphysical modality is about. This is something I will attempt to do without committing to any *particular* view of metaphysics, for reasons that will become clear.

Broadly speaking, the field of metaphysics, like the field of logic, concerns itself with specifying laws about what can or cannot be true. Since the laws of logic are universal, all metaphysical possibilities are also logical possibilities. However, the converse seems to be false: the laws of metaphysics seem to be less permissive than the laws of logic, meaning some logical possibilities are not metaphysical possibilities.

To get a sense of what the laws of metaphysics are, it helps to have a sense of what metaphysics is about. There are several characteristically metaphysical domains of inquiry, but one of the most obvious is ontology: the domain in which one tries to answer very general questions about existence and the types of things that exist. For example, the question 'Do other possible worlds exist?' is an ontological question. Since ontology is a metaphysical domain, any possible world that exists is *metaphysically* possible by default - at least, this is the case on the Lewisian conceptual scheme of possible worlds to which I am helping myself. That means that the set of metaphysically possible worlds is *all there is*.[5] So, to get a sense of what is metaphysically possible, one can just try to imagine what goes on in worlds that might exist.

In trying to imagine what goes on in worlds that might exist, we are of course guided by intuition, just as with the laws of logic; but thankfully, just as with the laws of logic, our intuitions once again seem to converge, at least some of the time. For example, since the identity of things is another characteristically metaphysical subject matter, it seems to be a metaphysical necessity that water is H_2O. As Kripke writes:

> We identified water originally by its characteristic feel, appearance and perhaps taste, (though the taste may usually be due to the impurities). If there were a substance, even actually, which had a completely different atomic structure from that of water, but resembled water in these respects, would we say that some water wasn't H_2O? I think not. We would say instead that just as there is a fool's gold there could be a fool's water; a substance which, though having the properties by which we originally identified water, would not in fact be water. And this, I think, applies not only to the actual world but even when we talk about counterfactual situations. (Kripke, 1980, p. 128)

If Kripke is right, then the laws of metaphysics are less permissive than the laws of logic, because logical necessities can be established *a priori*, whereas establishing that water is H_2O required empirical investigation. In other words, 'Water is H_2O' is an *a posteriori* necessity; hence, not a purely logical one.

[5] Predictably, there is nonetheless an interesting philosophical literature on metaphysically impossible worlds. See e.g. Yagisawa (1988) and Nolan (1997). I do not entertain this literature here.

All this talk of other worlds that might exist may seem mysterious and fantastical, especially given that, if they do exist, we are nonetheless causally isolated from them.[6] With this in mind, one might wonder how we are supposed to gain knowledge of the laws of metaphysics. But as Kripke says, '[g]enerally, things aren't 'found out' about a counterfactual situation... as if we were looking at them through a telescope.' (1980, p. 49) That is, we do not investigate other possible worlds and then identify their objects and events according to their characteristic properties. Rather, we *stipulate* which real world objects/events we are talking about and then imagine worlds in which those objects/events have their characteristic properties. For chemical compounds like water, chemical structure is one such characteristic property.

Of course, some will claim that this is the wrong way to think about the metaphysical modality, and the wrong way to explain the metaphysical necessity of the proposition expressed by 'Water is H_2O.' Others will claim that metaphysical necessities do not require explanation: they are just brute facts. Others again will claim that 'Water is H_2O' is *not* a metaphysical necessity. And others still will argue that there is simply no fact of the matter because metaphysics is dangerous nonsense. Such is the range of beliefs regarding the field of metaphysics, even among philosophers. Nonetheless, the above should suffice to give the reader a sense of what the metaphysical modality is about.

With this said, let us now consider the metaphysically strict analysis of indicative conditionals:

metaphysically strict analysis $A \rightarrow C$ is true iff $A \supset C$ is metaphysically necessary.

The first thing to point out when defending the metaphysically strict analysis is that it meets the minimal respect desideratum - that is, it does not make logically impossible any conditional *or* unconditional propositions that are commonly intuited as true by competent users of the language. To see this, first note that if the metaphysically strict analysis is correct, then the laws of logic do not suffice to determine the truth values of any indicative conditionals, because *the truth values of indicative conditionals depend in part on the laws of metaphysics*. Hence, the metaphysically strict analysis does not make any indicative conditionals *logically* impossible. Second, note that a theory of indicative conditionals makes other propositions (i.e. propositions other than indicative conditionals) logically impossible only indirectly, by way of making indicative conditionals either true or false. Hence, the metaphysically strict analysis does not make any other propositions logically impossible either. Whatever the intuitions of competent language

[6] Again, this is the case on the Lewisian view. Lewis's view should therefore be contrasted with e.g. the Everettian many-worlds view - see Barrett (2023, §5.3).

users, the metaphysically strict analysis of indicative conditionals meets the minimal respect desideratum.

The fact that the metaphysically strict analysis of indicative conditionals meets the minimal respect desideratum means that its conjunction with the laws of logic does not suffice to make truth value claim 3 false. But more than this, the metaphysically strict analysis plausibly makes truth value claim 3 *true*. Since Glasgow is a part of Scotland, $\mathbf{G} \supset \mathbf{S}$ seems to be necessary in some sense or other, and since mereology (i.e. the relations of parts to wholes) is another characteristically metaphysical subject matter, $\mathbf{G} \supset \mathbf{S}$ is plausibly *metaphysically* necessary in particular. Hence, on the metaphysically strict analysis, $\mathbf{G} \rightarrow \mathbf{S}$ is plausibly true. On the other hand, my actual place of birth seems to be metaphysically contingent: if I happened to exit the womb elsewhere in the United Kingdom, I would presumably still be *me*. So, despite being born in neither Glasgow nor London, it seems to be metaphysically possible that I was born in either. More importantly, it seems to be metaphysically possible that I was born in Glasgow and *not* in England, or that I was born in London and *not* in Scotland. Hence, on the metaphysically strict analysis, both $\mathbf{G} \rightarrow \mathbf{E}$ and $\mathbf{L} \rightarrow \mathbf{S}$ seem to be false.

Of course, the laws of metaphysics *might* rule out my being born in either Glasgow or London. Or, in the opposite direction, they might allow for my being born in Glasgow but *not* in Scotland. Determining what is metaphysically possible is no easy task, and there is, partly for this reason, a long tradition of scepticism regarding metaphysics, championed by philosophers like David Hume, the logical empiricists of the Vienna Circle, and W. V. O. Quine. However, I think the metaphysically strict analysis is respectable even if the study of metaphysics is not. The important point is that the metaphysically strict analysis is not a theory about metaphysics. Rather, it is a theory about conditionals. Scepticism about metaphysics should not prevent one from believing the metaphysically strict analysis. Rather, it should affect which conditionals (if any) one believes on the supposition that the metaphysically strict analysis is true.

Suppose, for example, that one thinks statements about metaphysics do not have truth values. In that case, conditionals do not have truth values either. But as we have seen, that is a position that some theorists already defend. Alternatively, suppose one thinks that the notion of metaphysical necessity simply reduces to truth: nothing is metaphysically possible except what is true at the actual world. This is what Curley and Walski call 'strict necessitarianism'.[7] Strict necessitarianism is not a popular view,

[7] It may be helpful to distinguish strict necessitarianism from some closely-related views: determinism, actualism, and other forms of necessitarianism. Whereas strict necessitarianism is the view that all truths are metaphysically necessitated, determinism is the view that all truths are necessitated by the laws of nature (see the following section) plus all past and current events. Actualism, on the other hand, is the view that there are necessarily no *mere possibilia*, i.e. non-actual objects. In a slogan: 'to be is to exist, and to exist is to be actual.' (Menzel, 2023a, §0) Actualism may follow from strict necessitarianism, but strict necessitarianism does not follow from actualism. Lastly, other forms of necessitarianism involve

but there is evidence that it has occasionally been endorsed: see e.g. Griffin (2012, p. 58), who argues 'that the textual evidence for attributing [strict] necessitarianism to Leibniz and Spinoza is pretty strong', and see also Garrett (1991) for an influential defense of a strict necessitarian interpretation of Spinoza in particular. When held in conjunction with strict necessitarianism, the metaphysically strict analysis makes true all and only those indicative conditionals that have a false antecedent or a true consequent. But as we have seen, that is a position that some theorists already defend.

The fact that the metaphysically strict analysis mimics the material interpretation when paired with strict necessitarianism helps to explain what is appealing about the material interpretation: the material interpretation gets things right *if* strict necessitarianism is true. On the other hand, it also helps to explain what is unsatisfactory about the material interpretation: the material interpretation gets things right *only if* strict necessitarianism is true. And as Bennett (1996, §8) puts it, '[t]he view that this is the only possible world seems on the face of it to be tremendously implausible'.

None of the above is an attempt to persuade the reader of a particular theory of metaphysics. It is only to point out that the metaphysically strict analysis can be combined with different metaphysical theories, and different combinations yield different predictions, some of which match the predictions of existing theories. Far from giving us reason to doubt the metaphysically strict analysis, the wide array of existing metaphysical views actually helps to explain some of the appeal (and lack thereof) of existing theories. I think this is a virtue of the metaphysically strict analysis.

Of course, the metaphysically strict analysis also has its vices, the main one being that there are many intuitively true conditionals that it seems to make false (note: false, not logically impossible). For example, with the thought experiment from section 2.1 in mind, 'If Portugal did not win the game, then Qatar won it' ($\neg \mathbf{P} \rightarrow \mathbf{Q}$) is intuitively true even though 'Either Portugal won the game or Qatar won it' ($\neg \mathbf{P} \supset \mathbf{Q}$) does not seem to be metaphysically necessary. Mounting a full defence against this objection will be the purpose of chapter 4. What I have tried to do in the latter half of this section, and what I will continue to try to do, is show that the metaphysically strict analysis of indicative conditionals has some virtues. In the next section, I compare it with a similarly virtuous view: the nomically strict analysis. In the penultimate section, I argue that the metaphysically strict analysis is preferable given a plausible view of the laws of nature.

the claim that something other than truth is metaphysically necessary, e.g. necessitarians regarding the laws of nature believe that the laws of nature are metaphysically necessary. (See the following section.)

3.5 The Nomically Strict Analysis

Philosophers and laypeople alike often talk about the 'laws of nature', i.e. laws that govern the natural world. One might reasonably think that there is a type of possibility or modality arising from such laws. In philosophical parlance, this modality goes by many names: the 'natural' modality, the 'nomological' modality, or the 'nomic' modality, the latter being my nomenclature of choice.

The laws of nature are typically construed to include the laws of natural sciences such as physics, chemistry, and biology. Importantly, they govern *causal* interactions. Here is Fine on the matter:

> Natural necessity is the form of necessity that pertains to natural phenomena. Suppose that one billiard ball hits another. We are then inclined to think that it is no mere accident that the second billiard ball moves. Given certain antecedent conditions and given the movement of the first ball, the second ball *must* move. And the 'must' here is the *must* of natural necessity. (Fine, 2002, §2)

This example involving billiard balls is a paradigm case of causation, and the causal interaction between the billiard balls can be modelled using the laws of physics.

Additionally, the laws of nature are typically construed to be less permissive than the laws of metaphysics. As Kment writes, 'It is often assumed that nomic necessity is a weaker form of necessity than metaphysical necessity... so that anything that is metaphysically necessary is also nomically necessary, but not vice versa.' (2021, §2) Let us make the same assumption.

With this said, let us now consider the nomically strict analysis:

> **nomically strict analysis** $A \to C$ is true iff $A \supset C$ is nomically necessary.

Some of the early defenders of strict analyses seem to have had something like the nomically strict analysis in mind. Consider e.g. the following passage from Goodman:

> When we say
>
> > If that match had been scratched, it would have lighted,
>
> we mean that conditions are such, i.e. the match is well made, is dry enough, oxygen enough is present, etc., that "That match lights" can be inferred from "That match is scratched"... The principle that permits [this] inference... is not a law of logic but what we call a natural or physical or causal law. (Goodman, 1947, p. 116)

Even though the conditional in this passage is a counterfactual, it should be remembered that Goodman believes counterfactuals can be contraposed to become conditionals with true antecedents (see section 2.5). Hence, in his eyes, '[t]he problem of counterfactuals is equally a problem of factual conditionals'. (1947, p. 114)

Since the metaphysical necessity of $A \supset C$ implies the nomic necessity of $A \supset C$, some of the virtues of the metaphysically strict analysis are enjoyed also by the nomically strict analysis. For example, just as the metaphysically strict analysis can invoke the possibility of strict necessitarianism to help explain the appeal of the material interpretation, so the nomically strict analysis can do the same. In fact, some philosophers are necessitarians with regard to the laws of nature - see e.g. Swoyer (1982), Shoemaker (1980, 1998), Fales (1993), Ellis (2001), and Bird (2005). That is, they believe that the laws of nature are metaphysically necessary.[8] To such philosophers, there cannot be any difference in consequence between the nomically strict analysis and the metaphysically strict analysis. To compare the two views, then, we must consider the consequences of the nomically strict analysis on the assumption that the laws of nature are metaphysically contingent, as per e.g. Lewis:

> The worlds are many and varied. There are enough of them to afford worlds where... the physical constants do not permit life, or totally different laws govern the doings of alien particles with alien properties. (D. K. Lewis, 1986a, p. 2)

Like the metaphysically strict analysis, the nomically strict analysis meets the minimal respect desideratum, for reasons exactly analogous to before. If the nomically strict analysis is correct, then the laws of logic do not suffice to make any indicative conditionals true or false, because the truth values of indicative conditionals depend in part on the laws of nature. And since a theory of indicative conditionals makes other propositions logically impossible only by way of making indicative conditionals true or false, the nomically strict analysis does not make any propositions logically impossible.

Additionally, the nomically strict analysis makes true many intuitively true conditionals that seem to be false on the metaphysically strict analysis. For instance, with Fine's billiard balls in mind, the conditional 'If the first ball hits the second ball, then the second ball will move' seems to be true. On the nomically strict analysis, this is easy to explain, because the corresponding material conditional seems to be necessitated by the laws of nature. Of course, if necessitarianism regarding the laws of nature is correct, then the metaphysically strict analysis can also explain this; but necessitarianism regarding the laws of nature may not be correct. Hence, the nomically strict analysis has

[8] This kind of necessitarianism is defined in slightly different ways by different authors, many of whom put it in terms of *dispositions*. For example, Gozzano defines it as 'the view that dispositions, when stimulated, necessitate their manifestations.' (2020, p. 1).

an advantage over the metaphysically strict analysis in that, if contingentism regarding the laws of nature is correct, then the nomically strict analysis makes many more intuitively true conditionals true.

Despite this, the nomically strict analysis still makes many intuitively true conditionals false. For example, with the thought experiment from section 2.1 in mind, 'If Portugal did not win the game, then Qatar won it' ($\neg\mathbf{P} \rightarrow \mathbf{Q}$) is intuitively true even though 'Either Portugal won the game or Qatar won it' ($\neg\mathbf{P} \supset \mathbf{Q}$) does not seem to be necessitated by the laws of nature - if it is necessitated at all, it is necessitated by the rules of the game. Hence, even if one endorses the nomically strict analysis over the metaphysically strict analysis, one still has to explain the falsehood of a great many intuitively true conditionals. Explaining this widespread error will be the subject of the next chapter. In the next section, I argue that the nomically strict analysis faces an objection that the metaphysically strict analysis can easily avoid.

3.6 A Reason to Prefer the Metaphysically Strict Analysis

Alongside Lewis, I endorse a Humean view of the laws of nature known as the 'best-system analysis'. According to this view, something 'is a law iff it is a theorem of the best [deductive] system'. (1994, p. 478) A deductive system is simply a set of axioms that can act as premises in deductively valid arguments, thereby licensing inferences to conclusions, i.e. *theorems*. We want these theorems to be true (which they will be if the axioms are true), but we also want them to be informative - that is, we want them to give the deductive system explanatory and predictive strength. The best way to get a strong deductive system is simply to list all relevant true propositions as axioms. However, we also want a system to be *simple* - the fewer axioms the better, all other things being equal. The best system is that which 'strikes as good a balance as truth will allow between simplicity and strength.' (D. K. Lewis, 1994, p. 478)

Importantly, on this conception of the laws of nature, the laws are determined in part by what happens in the future, and this remains the case so long as there *is* a future. Additionally, they are relative to each world. For example, the laws at the actual world are determined by the history/future of the actual world, not the history/future of other possible worlds.

For anyone who endorses a Humean best-system analysis of the laws of nature, the nomically strict analysis gives rise to the following problem: the facts of the future help to determine the present history-to-future conditional truths, and the present history-to-future conditional truths help to determine the facts of the future. In other words, true history-to-future conditionals pull themselves up by the bootstraps. Suppose, for example, that $H \rightarrow F$ is true, where H is the actual world's history and F is a true proposition about a particular future event or state of affairs. Since F helps to determine the nomic necessity of $H \supset F$, it also helps to determine the present truth of $H \rightarrow F$. Yet the present truth of H and $H \rightarrow F$ implies the truth of F, and $H \rightarrow F$ thereby helps to determine that F is a fact of the future.

The problem is specific to future facts. Suppose P is a proposition about a past or present event or state of affairs. On the best-system analysis, P helps to determine the nomic necessity of $H \supset P$, and the present truth of H and $H \rightarrow P$ implies the truth of P. But intuitively, $H \rightarrow P$ does not thereby help to *determine* the relevant facts - that would seem to require that the conditional is true at a time *preceding* the occurrence of the relevant event or state of affairs. Thus, the problem arises on the Humean view because the facts of the future help to determine the laws of the present.

One way to describe the relationship sketched is as a relationship of mutual dependence. Of course, sometimes mutual dependence is unproblematic. For example, the amount of stock ordered by a grocery shop depends in part on the amount of custom it receives, and the amount of custom it receives depends in part on the amount of stock it orders. There is nothing problematic about this. However, the dependence in this case is between different instances of the things in question: the amount of stock ordered at time t_1 depends partly on the amount of custom received between t_0 and t_1, and this depends partly on the amount of stock ordered at t_0, but it does not depend on the amount of stock ordered at t_1. The dependence in the above case is different: it holds between the very same instances of the things in question, making it a case of vicious circularity.

One might wonder if the problem is really to do with assigning truth values in the present to propositions that are solely about the future. But the circularity does not depend on F's being true *in the present* - its being true in the future will do. What matters is whether the conditional $H \rightarrow F$ is true in the present, and on the nomically strict analysis, this just depends on whether there are any nomically possible worlds in which H is false or F is true *at some time or other*. Thus, whenever F becomes true, it will have helped to determine its own truth.

Of course, one might simply reject the Humean best-system analysis of the laws of nature - I do not pretend that the above is a knock-down argument against the nomically strict analysis. Nonetheless, *if* one conceives of the laws of nature as per the Humean best-system analysis, then one has to respond to the above objection in order to defend the nomically strict analysis. On the other hand, the above objection is easily avoided on the metaphysically strict analysis. On a popular view of the metaphysical and nomic modalities, they are fundamentally distinct from each other - see e.g. Fine (2002). Hence, even if one endorses a best-system analysis of the laws of nature, one need not endorse the same analysis of the laws of metaphysics. To my mind, the universality of the laws of metaphysics suggests that theories of metaphysics are more akin to theories of logic than to theories of the natural sciences. Like theories of logic, I take theories of metaphysics to be normative proposals that aim to respect our intuitions, where the relevant intuitions are about matters like identity and existence - matters that are logical in a *broad* sense, as Fine (2002, §1) puts it. Importantly, on this view, the laws of metaphysics need not be determined even in part by the facts of the future.

I do not hope to persuade the reader of any particular analyses of the laws of metaphysics or the laws of nature. Nonetheless, I do hope to persuade the reader that *if* one conceives of the latter (and only the latter) as per the Humean best-system analysis, then the metaphysically strict analysis is preferable to the nomically strict analysis. The main cost of the metaphysically strict analysis compared to the nomically strict analysis is that one has to explain the falsehood of a larger set of intuitively true conditionals. But since one has to explain the falsehood of a very large set in either case, the larger set seems to be a price worth paying in order to avoid the above objection.

There are at least three other reasons to prefer the metaphysically strict analysis that are worth mentioning. The first is hinted at by the following passage from Lewis's *On the Plurality of Worlds*:

> Sometimes one hears a short list of the restricted modalities: nomological, historical, epistemic, deontic, maybe one or two more. And sometimes one is expected to take a position, once and for all, about what is or isn't possible... I would suggest instead that the restricting of modalities... like the restricting of quantifiers generally, is a very fluid sort of affair: inconstant, somewhat indeterminate, and subject to instant change in response to contextual pressures. Not anything goes, but a great deal does. (D. K. Lewis, 1986a, p. 8)

By 'restricted modalities', Lewis means any modality whose domain is a proper subset of metaphysically possible worlds. Defining such domains in a *determinate* way is difficult, whereas defining the domain of metaphysically possible worlds in a determinate way is easy: it is simply the domain of all worlds that exist.

The other two reasons will become clear in chapters 6 and 7. In chapter 6, I argue that the metaphysically strict analysis can be extended beyond indicative conditionals to form a unified analysis of natural language conditionals more generally. That argument hinges on the claim that the laws of metaphysics are the same in all metaphysically possible worlds. I think this is a more plausible claim than the analogous claim that the laws of *nature* are the same in all *nomically* possible worlds. To put it another way: it is highly implausible that there are metaphysically possible worlds that have different laws of metaphysics, but it is not highly implausible that there are nomically possible worlds that have different laws of nature. In chapter 7, I argue in favour of a connection between conditionals and conditional probability such that 'If A, then C' is true iff the conditional probability of C given A is 1. The defence of this connection on the metaphysically strict analysis involves the claim that a proposition is metaphysically impossible only if its objective chance of occurrence is 0. I think this is a more plausible claim than the analogous claim that a proposition is *nomically* impossible only if its objective chance of occurrence is 0. To put it another way: it is highly implausible that some metaphysical impossibilities have a non-zero chance of occurring, but it is not highly implausible that some nomic impossibilities have a non-zero chance of occurring.

Having said all this, the arguments that I make in chapters 4 and 5 can be used to defend the nomically strict analysis as well as the metaphysically strict analysis, and it may be that similar arguments to those that are made in defense of the metaphysically strict analysis in chapters 6 and 7 can also be made in defense of the nomically strict analysis. To my mind, both views are worthy of consideration, and either is an improvement on the material interpretation.

3.7 Conclusion

In this chapter, I have explained and motivated what I call the 'minimal respect desideratum'. I have argued that the material interpretation and the epistemically strict analysis of indicative conditionals both fail to meet this desideratum, whereas the metaphysically strict analysis and the nomically strict analysis do not. Lastly, I have argued that the metaphysically strict analysis is preferable to the nomically strict analysis given a Humean best-system analysis of the laws of nature.

Even if one joins me in thinking that the metaphysically strict analysis is preferable to the nomically strict analysis, there is still a serious objection that needs to be responded to: if the metaphysically strict analysis is correct, then it seems many of the conditionals that we actually assert are false, including (no doubt) some of those that are asserted in this thesis. Defending against this objection will be the purpose of the next chapter.

Chapter 4

The Assertability of Conditionals

People say false things all the time. For example, the proposition expressed by the previous sentence is false, since people only really say false things *some* of the time. A more charitable interpreter may disagree, saying instead that I had a restricted domain of time in mind - restricted just enough to make the proposition true. This strategy for interpreting quantified statements allows statements like 'all the beer is in the fridge' to express true propositions while 'ignoring most of all the beer there is', to borrow an example from Lewis. (1986a, p. 3) But the charitable interpreter should also be open to the possibility that I just said something false. Sometimes, a false statement is as good as a true one.

In this chapter, I defend the metaphysically strict analysis of indicative conditionals against the objection that it makes too many intuitively true conditionals false. To do this, I invoke the widely recognised distinction between *semantics* (the domain of propositions and truth values) and *pragmatics* (the domain 'of linguistic acts and the contexts in which they are performed,' as Stalnaker (1970a, p. 275) puts it). Theories of semantics are often supported by theories of pragmatics. A well-known example is Grice's (1967b) theory of conversational implicature, which supports the material interpretation of indicative conditionals against the objection that it makes too many intuitively false conditionals true. The material interpretation has the consequence that having a false antecedent or a true consequent suffices to make an indicative conditional true, but Grice's theory of conversational implicature explains why one generally ought not to assert an indicative conditional just on the basis of a false antecedent or a true consequent. More importantly for our purposes, it explains why we intuitively judge some such conditionals to be false: we confuse the semantic property of falsehood with the pragmatic property of *unassertability*.

The metaphysically strict analysis of indicative conditionals has the opposite problem: it makes too many intuitively true conditionals false. Nonetheless, pragmatics can help: just as we sometimes take propositions to be false when they are merely unassertable, so (I suggest) we sometimes take propositions to be true when they are merely assertable. This is a controversial claim, because many philosophers endorse norms of assertion according to which assertability implies truth. For example, Williamson (1996) endorses the knowledge norm of assertion, according to which P is assertable only if one *knows* that P.[1] Nonetheless, I argue that one can coherently endorse the knowledge norm as a norm for the assertion of *un*conditional propositions while defending a separate norm for the assertion of indicative conditionals. In particular, I argue that one may assert an indicative conditional only if one knows that its corresponding material conditional is true in all possible worlds that are of interest to the participants in the conversation.[2] Importantly, I defend this as a *complete* account of the internal restrictions on the assertion of indicative conditionals, meaning some conditionals are assertable even though their corresponding material conditionals are false in some metaphysically possible worlds.

I begin by explaining Grice's theory of conversational implicature and the support that it lends to the material interpretation of indicative conditionals. I then introduce some key concepts in the literature on norms of assertion and explain some of the reasons given in favour of the knowledge norm. In section 4.3, I explain and defend a closely related Stalnakerian view of assertion according to which assertion is a way of narrowing down the set of live possibilities given a group's common knowledge. This leads me to argue in favour of the above norm of assertion for indicative conditionals, which in turn makes room for an explanation of why we intuitively judge some false conditionals to be true: we confuse the semantic property of truth with the pragmatic property of *assertability*.

4.1 The Gricean Defence of the Material Interpretation

Implicature is a phenomenon we see all the time: 'She turned around and walked away' *loosely* implies (or *implicates*) that first she turned around and then she walked away, as opposed to the other way round. Conversational implicature in particular is the kind of loose implication that arises from the presumption that one is following implicit norms of conversational conduct. Grice (1967b, pp. 26-7) expresses these norms as maxims categorised under the headings of Quantity, Quality, Relation, and Manner.

[1] More obviously, Dummett (1973) and Weiner (2005) endorse the truth norm of assertion, according to which P is assertable only if P is true. Since this is logically weaker than the knowledge norm, I focus my attention on the latter.

[2] Of course, proponents of the knowledge norm may agree with this norm of assertion *if* they endorse some other view of indicative conditionals besides the metaphysically strict analysis. What I argue is that *even if* one endorses the metaphysically strict analysis, one can motivate this norm of assertion for indicative conditionals, and one can do so in a way that simultaneously motivates the knowledge norm as a norm for the assertion of unconditional propositions.

For example, under the category of Quality, he writes 'Do not say what you believe to be false', and 'Do not say that for which you lack adequate evidence'. (1967b, p. 27)

Implicit in some of Grice's maxims is the supposition that propositions have relevant alternative propositions that might be asserted instead. For example, the first maxim of Quantity, 'Make your contribution as informative as is required' (1967b, p. 26), prohibits one from asserting a proposition if there is a more informative relevant alternative proposition that one may assert in that context instead. Similarly, the third maxim of Manner, 'Be brief' (1967b, p. 27), prohibits one from asserting a proposition if there is a *shorter* relevant alternative proposition that one may assert in that context instead.[3]

As we saw in section 2.1, all material conditionals with false antecedents are true, and all material conditionals with true consequents are true. However, on the supposition that indicative conditionals are material, Grice (1967a) uses the theory of conversational implicature to reach the conclusion that, in general, one ought not to assert an indicative conditional solely on either of these bases. If we suppose that indicative conditionals are material, then both $\neg A$ and C are relevant alternative propositions to $A \rightarrow C$, since they both entail $A \rightarrow C$. But by the maxim of Manner mentioned above, one ought not to assert $A \rightarrow C$ if there is a shorter relevant alternative proposition that one may assert instead. Indeed, since both $\neg A$ and C entail $A \rightarrow C$, they are shorter *and* more informative. So whenever one may assert either $\neg A$ or C, the indicative conditional $A \rightarrow C$ is unassertable.

This Gricean route to unassertability helps the proponent of the material interpretation to explain why we intuitively judge some conditionals with false antecedents or true consequents to be false: we mistake their unassertability for falsehood. The idea features also in the writing of Jackson:

> Consider the conditional 'if Jones lives in London, then he lives in Scotland.' It is plausible that we can say straight off that this conditional is false without waiting to find out whether or not Jones lives in London, or whether or not he lives in Scotland... The only reply to this objection that is at all plausible is to argue that all we can say straight off about 'if Jones lives in London, then he lives in Scotland' is that it is highly unassertable, and that this is compatible with the conditional turning out to be true... The immediately evident property of the conditional is high unassertability, not falsity. (Jackson, 1991, p. 2)

The reply is not just 'at all plausible', but *highly* plausible, since the lack of consensus regarding the truth conditions of conditionals makes it very easy for us to mistake conditionals as false when they are merely unassertable.

[3] This concept of relevant alternative proposition is not to be confused with the concept of relevant alternative employed in contextualist epistemology by e.g. Dretske (1970) and Goldman (1976).

An analogous defence of the metaphysically strict analysis can be made by claiming that some conditionals are mistaken as true when they are merely assertable. The strongest reason to reject this defence is just that assertability seems to *imply* truth. In the next section, I explain a norm of assertion according to which the assertability of *P* implies not just that *P* is true, but that it is *known*.

4.2 The Knowledge Norm of Assertion

Assertion is a type of speech act performed by saying a declarative sentence. For example, in most contexts, if I say to my friend 'Boris Johnson has resigned as a Member of Parliament,' I thereby *assert* that Boris Johnson has resigned as a Member of Parliament. Some contexts are exceptional: if, for example, I say the same sentence while reading aloud from a newspaper, I have not thereby asserted that Boris Johnson has resigned as a Member of Parliament; likewise if I say the same sentence as part of a theatre production. Asserting is therefore more specific than saying. Nonetheless, we have an intuitive grasp of it, and that intuitive grasp can be clarified through philosophical investigation.

The literature on norms of assertion has grown considerably within the last few decades, thanks in large part to the influence of Williamson. In his (1996) 'Knowing and Asserting', Williamson suggests that assertion is governed not just by norms pertaining to conversation in general à la Grice, but by a constitutive norm that is specific to the practice of assertion itself. As he explains, constitutive norms are like the rules of a game: if one changes the rules of a game, then one creates a new game; and if one tries to create a game with exactly the same rules as an existing game, then one *fails* to create a new game. Hence, if there are any constitutive norms of assertion, they govern assertion necessarily and uniquely. This makes norms of assertion distinct from e.g. moral norms, which govern other types of action as well.

The particular constitutive norm of assertion argued for by Williamson is the knowledge norm of assertion, which he expresses as 'One *must*: assert that *P* only if one knows that *P*.' (1996, p. 494, my italics) This expression of the knowledge norm features a deontic "necessity" operator in wide scope, but other expressions of the knowledge norm feature a deontic "possibility" operator in the antecedent, as in 'One *may* assert that *P* only if one knows that *P*' - see e.g. McKinnon (2015, §1). To avoid issues of deontic logic, let us opt for an expression that does away with the deontic operator. In general, when an action conforms to a norm, we say that it is 'proper'.[4] With this in mind, let us express the knowledge norm of assertion as follows:

[4] For example, according to a norm of dinner table etiquette, the proper way to set a table is with the knife on the right hand side of the plate and the fork on the left hand side.

KNA One's assertion that *P* is proper only if one *knows* that *P*.[5]

KNA is what Williamson (1996, pp. 492-3) calls a *simple* account of assertability, meaning in part that it specifies a single restriction on assertion relative to the content of the assertion. It is also typically advocated as a complete account of the *internal* restrictions on assertion. If there were no external restrictions, then it would express necessary and sufficient conditions; as it is, it only expresses a necessary condition.

KNA has explanatory power. For example, it can explain the intuition that there is something infelicitous about an assertion of the form '*P* and I do not know that *P*': if I make an assertion of this form while conforming to KNA, then I know that *P and* I do not know that *P*. Contradiction. However, if KNA governs the assertion of conditional propositions, then false-but-assertable conditionals are impossible, since knowledge is factive. KNA thereby threatens to undermine my proposed defence of the metaphysically strict analysis of indicative conditionals.

There are many ways one might try to respond to this threat. Most obviously, one might simply argue for an alternative norm. Here are some alternatives to KNA, none of which make false-but-assertable conditionals impossible:

BNA One's assertion that *P* is proper only if one *believes* that *P*.

RBNA One's assertion that *P* is proper only if one *reasonably believes* that *P*.[6]

RTBNA One's assertion that *P* is proper only if it is *reasonable* (for one) *to believe* that *P*.[7]

Like KNA, these norms are simple accounts of assertability. At least some of them are implied by KNA (since knowledge implies justified belief), but they are typically advocated as *complete* accounts of the internal restrictions on assertion.

Despite the range of alternative norms that have been defended in the literature, my response to the threat posed by KNA is not to reject it outright. Rather, let us suppose that KNA governs the assertion of *un*conditional propositions only. This requires making an exception of conditional propositions, but that is not a novel idea. Back in section 2.6, we encountered a popular idea in the literature on conditionals known as:

[5] This expression of KNA presupposes that '...only if...' can be used to express a necessary condition (even outside the scope of a necessity operator). Strictly speaking, this is question-begging, since my overall aim is to *argue* that conditionals are modal in this way. However, the reader may rest assured that this assumption does not do any philosophical work - I express KNA this way only for the sake of clarity. And for what it is worth, this usage is perfectly commonplace. Indeed, '...if and only if...' (i.e. '...iff...') is our standard way of expressing *necessary* and *sufficient* conditions in analytic philosophy. It is grist to my mill that philosophers implicitly take conditionals to be modal when they use them this way.

[6] See Douven (2006) for a defence of a norm very much like this one.

[7] This last norm does not require that the asserter actually *does* believe that *P*. See Lackey (2007) for a prominent defence.

Adams' Thesis $\mathrm{As}(A \to C) = \mathrm{P}(C|A)$, except where $\mathrm{P}(A) = 0$.

What the thesis says is that the assertability of an indicative conditional is equal to the conditional probability of its consequent given its antecedent. On the face of it, this makes the assertability of indicative conditionals quite exceptional. Moreover, as Hájek writes, Adams' Thesis 'is taken by many to be a touchstone of any theorizing about indicative conditionals'. (2012a, p. 145)

Admittedly, Adams' use of the term 'assertability' is somewhat non-standard (see his (1965, 1975)), and Hájek tells us that, through 'personal communication' with Adams, he has been informed that 'what he had in mind involved reasonableness of belief more than appropriateness of utterance'. (2012a, p. 147) Nonetheless, the standard interpretation of Adams' Thesis does seem to be in terms of assertability of some sort, since this is the word that Adams used. As Lewis puts it:

> Ernest Adams has pointed out an apparent exception. In the case of ordinary indicative conditionals, it seems that assertability goes... by the conditional subjective probability of the consequent, given the antecedent. (D. K. Lewis, 1976, p. 297)

There are of course ways of making this apparent exception unexceptional. Back in section 2.5, we came across the idea that the probability of a conditional is equal to the conditional probability of its consequent given its antecedent:

Stalnaker's Thesis $\mathrm{P}(A > C) = \mathrm{P}(C|A)$, except where $\mathrm{P}(A) = 0$.

If one combines Stalnaker's Thesis with the idea that assertability is equal to probability, then Adams' Thesis is unexceptional. However, Stalnaker's Thesis has long been widely rejected, thanks in large part to the arguments presented by Lewis later on in his (1976), referenced above.[8]

To be clear, whichever way one interprets Adams' Thesis, I do not endorse it. I mention it here only to point out that there is nothing novel about the idea of making an exception of indicative conditionals when it comes to the matter of assertability. In the next section, I explain a view of assertion inspired by Stalnaker. This view lends support to KNA as a norm for the assertion of unconditional propositions, but when combined with the metaphysically strict analysis of indicative conditionals, it lends support to a separate norm for the assertion of indicative conditionals.

4.3 A Stalnakerian View of Assertion

One reason for thinking that there are constitutive norms of assertion is that the practice of assertion seems to have a characteristic *purpose* or *function*. As Simion writes:

[8] More on this in chapter 7.

> [W]hat epistemic goods is assertion meant to deliver? One very plausible answer is this: the characteristic purpose of assertion is generating testimonial knowledge in the hearer. (Simion, 2016, p. 3042)

Similarly, Kelp (2018) argues that we can model the practice of assertion on a simple economic system, and that the generation of knowledge is the etiological function of that system because it constitutes a benefit that contributes to the explanation of why individual acts of assertion occur.

I think Simion and Kelp are right that the generation of knowledge in the hearer is a characteristic purpose/function of assertion. However, things can have multiple characteristic purposes/functions. For example, the human mouth is used for eating, breathing, speaking, and much more. The same may be true of assertion. Individual acts of assertion are sometimes about *sharing* information rather than *trading* it, because there is value in having information in common with others. That is, sometimes we assert things to others because we want our knowledge sets to overlap. In line with this, I suggest that another characteristic purpose/function of assertion is an increase in *common* knowledge between asserter and hearer. This is almost always a consequence of generating testimonial knowledge in the hearer anyway.

Stalnaker has a view of assertion that coheres nicely with this idea. According to Stalnaker, assertion is a matter of reducing what he calls the 'context set', i.e. 'the set of possible worlds recognized by the speaker to be the "live options" relevant to the conversation.' (1999, pp. 84-5) He writes:

> To make an assertion is to reduce the context set in a particular way, provided that there are no objections from other participants in the conversation. The particular way in which the context set is reduced is that all of the possible situations incompatible with what is said are eliminated. (Stalnaker, 1999, p. 86)

By defining the context set as the set of worlds 'recognized' as live possibilities, Stalnaker introduces an element of subjectivity. Instead, let us think of assertion as a matter of narrowing down the set of live possibilities itself. This set of live possibilities should be construed as the set of metaphysically possible worlds that are logically consistent with the intersection of the participants' knowledge sets. Alternatively, one can think of it as the set of epistemic possibilities relative to a fictional epistemic agent: one who knows only those propositions that are known by all participants in the conversation.

Stalnaker goes on to draw an analogy with a game:

> One may think of a nondefective conversation as a game where the common context set is the playing field and the moves are either attempts to reduce the size of the set in certain ways or rejections of such moves by others. The

> participants have a common interest in reducing the size of the set, but their interests may diverge when it comes to the question of how it should be reduced. (Stalnaker, 1999, p. 88)

This is an extremely common analogy in the literature on assertion. For example, as we saw above, Williamson uses the rules of a game to explain the concept of constitutive norms. Similarly, Lackey compares the flouting of a norm to the breaking of a rule in a game: to do either is to be 'subject to criticism' for doing something 'improper'. (Lackey, 2007, p. 595) And in his (1973, p. 298), Dummett conceives of assertion as a game in which the end goal is saying something true.

Part of the appeal of Stalnaker's view of assertion is that it works especially well in the analogy with a game: rather than thinking of assertion as a game that an individual wins by asserting a true proposition, we can think of it as a *cooperative* game that participants win by reducing the set of live possibilities in the desired way. Another appealing aspect of Stalnaker's view is that it helps to justify KNA as a norm of assertion for unconditional propositions: if one wants to reduce the set of live possibilities in a particular way, then one ought to conform to KNA for the assertion of unconditional propositions. And if one believes that P without knowing that P, one can still assert that one believes that P so long as one knows that one believes it.

However, when held in conjunction with the metaphysically strict analysis of indicative conditionals, the Stalnakerian view leads us to expect that the assertion of indicative conditionals should be governed by something other than the knowledge norm. To see this, first note that the assertion of a metaphysically necessary proposition cannot directly reduce the set of live possibilities by generating testimonial knowledge of the content of the assertion. A metaphysically necessary proposition is true at all metaphysically possible worlds; *a fortiori*, it is true at all members of the set of live possibilities. Second, note that all metaphysically possible worlds are equivalent with regard to their metaphysical possibilities.[9] Now suppose that the metaphysically strict analysis of indicative conditionals is correct. In that case, an indicative conditional $A \rightarrow C$ is true iff $A \supset C$ is true at all metaphysically possible worlds; but if $A \supset C$ is true at all metaphysically possible worlds, then $A \rightarrow C$ is also true at all metaphysically possible worlds. Hence, a true indicative conditional is a metaphysical necessity. That means that the assertion of an indicative conditional cannot directly reduce the set of live possibilities by generating testimonial knowledge of the content of the assertion. Instead, it can only reduce the set of live possibilities indirectly, by licensing inferences to metaphysically *contingent* propositions.

[9] This claim is not indisputable, but it is a common assumption. To be thorough, I give an argument in its favour in chapter 6. For now, let us note that it is true if we assume a Lewisian conception of metaphysical possibility as truth at a world that exists. More generally, it is true if we assume that the modal logic system known as 'S5' correctly models metaphysical possibility.

All of this fits perfectly with Ryle's (1949, pp. 110-11) description of indicative conditionals as 'inference-tickets'. Indeed, given the above, we ought to expect the assertability conditions of indicative conditionals to track their ability to function as inference-tickets, since this is their way of contributing to the goal of narrowing down the set of live possibilities. And on the metaphysically strict analysis, an indicative conditional does not need to be true to function as an inference-ticket. If A is true at all worlds within a given set, then $A \rightarrow C$ functions as an inference-ticket to the truth of C (at those worlds) so long as $A \supset C$ is true throughout the set. Similarly, if C is false at all worlds within a given set, then $A \rightarrow C$ functions as an inference-ticket to the falsehood of A (at those worlds) so long as $A \supset C$ is true throughout the set. These are the distinctive ways in which an indicative conditional may be used as an inference-ticket, and an indicative conditional may be used these ways so long as its corresponding material conditional is true throughout the relevant set of worlds. Yet, if the relevant set of worlds is a proper subset of all metaphysically possible worlds, then $A \rightarrow C$ functions as an inference-ticket even if $A \supset C$ is not metaphysically necessary. Hence, on the metaphysically strict analysis, an indicative conditional does not need to be true to function as an inference-ticket. Bennett once wrote: 'Sometimes one does coherently get from falsehood to truth, but only through luck.' (2003, p. 84) The view that I am endorsing straightforwardly implies that he is wrong.

4.4 A Norm of Assertion for Conditionals

Let us use the term 'worlds of interest' to refer to those worlds about which the participants in a conversation want to make inferences. The following norm of assertion for indicative conditionals is sufficient to ensure that asserted conditionals function as inference-tickets, and that the practice of asserting conditionals thereby contributes to the goal of narrowing down the set of live possibilities:

> **KNAC** One's assertion that $A \rightarrow C$ is proper only if one knows that $A \supset C$ is true at all worlds of interest.[10]

Given that I am enlisting assertability conditions to do some of the work that truth

[10] Perhaps a weaker norm than KNAC is true, e.g.

> **BNAC** One's assertion that $A \rightarrow C$ is proper only if one *believes* that $A \supset C$ is true at all worlds of interest.
>
> **RBNAC** One's assertion that $A \rightarrow C$ is proper only if one *reasonably believes* that $A \supset C$ is true at all worlds of interest.
>
> **RTBNAC** One's assertion that $A \rightarrow C$ is proper only if it is *reasonable* (for one) *to believe* that $A \supset C$ is true at all worlds of interest.

However, if KNAC can help to show that an assertable conditional need not be true, then so can any of these other norms.

conditions are usually made to do, it should be no surprise that KNAC looks quite similar to the Lewisian and Stalnakerian truth conditions first encountered back in chapter 2. However, an important difference is that the worlds of interest need not be the most similar worlds to the actual world - sometimes, people want to make inferences about far off metaphysically possible worlds, such as those in which people are morally perfect.

Even though KNAC specifies only a necessary condition, like the simple accounts of assertability mentioned above, it should be read as a *complete* account of the relevant internal restrictions. Thus, we can infer that the assertion of an indicative conditional is proper *if* it meets the necessary condition specified in KNAC *unless* it flouts a more general conversational norm impacting assertability. One might object that the assertion of a false conditional is highly likely to flout one of the norms captured by Gricean maxims of Quality: 'Do not say what you believe to be false', and 'Do not say that for which you lack adequate evidence'. (1967b, p. 27) However, these maxims do not capture *more general* conversational norms - rather, they apply specifically to assertion. Indeed, it has even been suggested (by both Gazdar (1979, p. 46) and Rieger (2006)) that we replace these maxims with a knowledge-based maxim capturing the knowledge norm of assertion: '[S]ay only that which you know.' (Rieger, 2006, p. 235)

Since KNAC is to be read as a complete account of the internal restrictions on the proper assertion of conditionals, combining KNAC with the metaphysically strict analysis makes room for the possibility that many false indicative conditionals are nonetheless assertable. This in turn opens the door to the aforementioned defence of the metaphysically strict analysis: many of the conditionals that are rendered false by the metaphysically strict analysis are mistaken as true when they are merely assertable.

Indeed, if KNAC is true, then a great many indicative conditionals are assertable. One might even worry that *too* many indicative conditionals are assertable. Suppose one knows that $A \supset C$ is true at all worlds of interest just on the basis that $\neg A$ is true at all such worlds. Or suppose one knows that $A \supset C$ is true at all worlds of interest just on the basis that C is true at all such worlds. In either case, the assertion of $A \rightarrow C$ meets the condition specified in KNAC. However, I do not think this is anything to worry about. First of all, some conditionals clearly are assertable on these bases: Dutchman conditionals (e.g. 'If the butler is the murderer, then I'm a monkey's uncle') exploit assertability on the basis of a false antecedent, whereas 'even if'-conditionals (e.g. 'Even if you call a tail a leg, a dog still has four legs') exploit assertability on the basis of a true consequent. These exceptions aside, it is true that people tend not to assert indicative conditionals on either of these bases, but we can explain this linguistic data by appeal to the fact that indicative conditionals can only be used as inference-tickets in their distinctive ways if one can either *affirm* the antecedent or *deny* the consequent.

If one knows that $\neg A$ is true at all worlds of interest or that C is true at all worlds of interest, then $A \rightarrow C$ cannot be used as an inference-ticket to the truth of either of these propositions at all worlds of interest. In most cases, this is sufficient to make its assertion pointless.

4.5 Conclusion

In this chapter, I have argued that the metaphysically strict analysis of indicative conditionals can defend itself against the objection that it makes too many intuitively true conditionals false. In particular, I have argued that it can defend itself by appeal to the possibility that we sometimes mistake false conditionals as true because they are *assertable*. I began by explaining Grice's theory of conversational implicature and his defence of the material interpretation of indicative conditionals from which my defence of the metaphysically strict analysis draws inspiration. I then explained why the popularity of KNA makes this defence seem at first to be a non-starter. However, far from being a non-starter, the defence turns out to be quite defensible. In particular, it can be motivated by a Stalnakerian view of assertion - a view which also helps to motivate KNA as a norm for the assertion of unconditional propositions.

Over the next two chapters, I make a case for extending the metaphysically strict analysis to counterfactual conditionals. That case is made more plausible by KNAC. In section 5.7, on the supposition that corresponding indicative and counterfactual conditionals are logically equivalent, I show that KNAC can explain why some corresponding indicative and counterfactual conditionals nonetheless seem to differ in terms of their assertability.

Chapter 5

Corresponding Conditionals

When it comes to the analysis of natural language conditionals, an important question to ask is 'How many different logical types of natural language conditional are there?' I believe the answer to this question is 'one', meaning I endorse a monist or *unified* theory of conditionals.[1] However, most analytic philosophers who have written on the subject endorse a dualist theory, meaning they believe the answer is 'two'.[2] One way to support the view that there are two logical types of natural language conditional is to appeal to a pair of corresponding conditionals, i.e. conditionals that have the same antecedent and consequent as each other. The most well-known example of this is found in the opening pages of Lewis's *Counterfactuals* (1973, p. 3), where he appeals to the following conditionals:

(1) *If Oswald did not kill Kennedy, then someone else did*.

(2) *If Oswald had not killed Kennedy, then someone else would have*.

The general idea is that (1) and (2) seem to be logically nonequivalent (indeed, they seem to have different truth values), yet their constituent propositions seem to be the same (even though the words used to express them are different). Therefore, there must be two logical types of conditional. Or so the argument goes.

Given that (1) and (2) are indicative and counterfactual conditionals respectively, they are usually used to defend a version of dualism that draws the logical division between indicative and counterfactual conditionals in particular. The example is borrowed from Adams (1970), and it has been appealed to by many others, including Davis (1979), Edgington (1995), and Dudman (2000). Some (e.g. Bennett (1988) and Jackson (1990)) have used an analogous example involving Booth and Lincoln, whereas others (e.g. Edgington (1995)) have pointed to entirely different examples that seem to support drawing the logical division in the same place.

[1] Other proponents of this sort of view include Stalnaker (1968), Ellis (1978), and Starr (2014).

[2] E.g. Adams (1970), Lewis (1973), Davis (1979), Jackson (1990), Edgington (1995), Bennett (2003), and many more.

At the same time, analogous arguments have occasionally been used to defend the thesis that corresponding indicative and counterfactual conditionals are logically *equivalent*. For example, in two brief papers published shortly after Lewis's *Counterfactuals*, E. J. Lowe argues against Lewis by pointing out that (2) seems to have a logically equivalent indicative counterpart:

(3) *If Oswald* has *not killed Kennedy, then someone else* will *have*.

In this chapter, I argue that all such arguments from correspondence are unsatisfactory. The upshot of this, I suggest, is that we should look beyond such arguments when trying to conclusively answer the question 'How many different logical types of natural language conditional are there?' I begin by supposing the truth of dualism and explaining the meanings of 'indicative', 'counterfactual', and other terms used by dualists to distinguish the two logical types of conditional. I show that none of these terms are fully satisfactory as labels for the two logical types. I then reconstruct Lewis's argument for dualism, suggesting that it ought not to be interpreted as an argument for the traditional version of dualism according to which indicative and counterfactual conditionals are logically nonequivalent; rather, it ought simply to be interpreted as an argument for dualism of some version or other. I then explain Lowe's attack on Lewis's argument, which makes sense only if we interpret Lewis as arguing for the traditional version of dualism. This leads me to draw a distinction between what I call arguments from 'natural correspondence' and arguments from 'traditional correspondence'. I then iron out some confusions in Lowe's position before arguing that neither type of argument from correspondence is satisfactory: arguments from traditional correspondence involve extremely weak inferences, and arguments from natural correspondence either involve extremely weak inferences or appeal to beliefs that are not common ground. Before concluding, I offer some thoughts in favour of the idea that corresponding indicative and counterfactual conditionals are logically equivalent, and I show that KNAC (the norm of assertion for indicative conditionals defended in chapter 4) can explain our different reactions to (1) and (2) *if* monism is true.

5.1 Categorising Conditionals

Suppose dualism about conditionals is true, meaning there are two logical types of conditional. As Bennett writes, '[i]t would be good to label each in a way that helpfully describes it, rather than being stipulated and meaningless ('Type One', 'Type Two'), or meaningful and false.' (2003, §5) In other words, it would be good to identify some non-logical features in virtue of which the two types of conditional can be distinguished from each other. 'Unfortunately,' he continues, 'we have no labels with this virtue.' (2003,

§5) In what follows, I explain and evaluate the labels that are in common usage. Much of what I say in this section has already been said (see e.g. Edgington (1995, §1.2) and Bennett (2003, §5)), but saying it again here will help the reader to make sense of the following sections, the content of which has *not* been said before.

Lewis uses the terms 'indicative' and 'counterfactual' (1973, p. 3) to refer to what seem to be the distinguishing features of the two conditionals mentioned above. While both of these terms refer to non-logical properties with which we are already familiar, it is generally agreed that they do not identify properties in virtue of which conditionals can be assigned to one logical category or the other. In other words, they are 'meaningful and false' as opposed to 'stipulated and meaningless'. To explain why this is, I will use the labels dismissed by Bennett: 'Type One' for conditionals of the same logical type as (1), and 'Type Two' for conditionals of the same logical type as (2).

Let us begin with Lewis's preferred label for Type Two conditionals: 'counterfactuals'.[3] On the face of it, this label suggests something about the truth value of a conditional's antecedent: in particular, it suggests that it is false. Accordingly, one might insist that a conditional is genuinely counterfactual iff its antecedent is false. However, by that definition, both (1) *and* (2) are counterfactual, so this cannot be what distinguishes one conditional from the other. More generally, since corresponding conditionals are supposed to have the very same antecedent as each other, the antecedent's truth value cannot distinguish one conditional from the other. Instead, a better way to interpret 'counterfactual' in this context is as follows: a conditional is counterfactual iff it conveys a belief in the antecedent's negation. By this definition, (2) is a counterfactual, and (1) is not, whether or not Oswald killed Kennedy.

Being counterfactual in this sense may distinguish (2) from (1), but it does not distinguish all Type Two conditionals from all Type One conditionals. Consider the following dialogue:

Holmes: *There is blood on the butler's uniform, and his fingerprints are on the murder weapon.*

Watson: *If the butler were the murderer, then that is exactly what we would expect to find!*

Watson's conditional does not convey a belief in the antecedent's negation - it is being asserted as a reason for thinking that the butler *is* the murderer, so the speaker is clearly open to believing the antecedent. Hence, some Type Two conditionals are not counterfactual by the above definition.

A better definition might be that counterfactuals just convey a lack of belief in the antecedent as opposed to a belief in the antecedent's negation. By this definition,

[3] One also sometimes sees the synonymous labels 'contrafactual' (e.g. Quine (1950, p. 21)) and 'contrary-to-fact' (e.g. Chisholm (1946)).

Watson's conditional is plausibly counterfactual. But so are some Type One conditionals. Consider e.g. 'If the butler is the murderer, then I'm a monkey's uncle'. The whole purpose of Dutchman conditionals like this one is to express doubt regarding the conditional's antecedent.

Perhaps there is some better label for Type Two conditionals, but I think not. The only alternative to 'counterfactual' in the literature is 'subjunctive', which refers to the grammatical mood in which Type Two conditionals are purportedly expressed. Unfortunately, definitions of the subjunctive tend to be somewhat vague - as Edgington points out, 'Grammarians are no more prone to unanimity than philosophers'. (1995, p. 240) Roughly, the idea is that the subjunctive mood is characterised in English by the plain form of the verb (i.e. the form that features in the infinitive), as in archaic phrases like 'Praise *be* to God', and formal language like 'It is crucial that he *arrive* early.' As such, the subjunctive is normally used to make statements about unactualised possibilities. Type Two conditionals are also usually about unactualised possibilities, but they tend not to make use of the plain form of the verb - indeed, the subjunctive mood has all but died out in modern English.

Might we justify the 'subjunctive' label some other way? Some sources teach that the word 'were', which often features in the antecedent of Type Two conditionals, is the past tense version of the present subjunctive 'be'. However, there is reason to doubt this, since one cannot say things like 'Praise *were* to God'.[4] Anyway, there are many Type Two conditionals that do not feature the word 'were', for example: 'If Oswald had not killed Kennedy, then someone else would have.'

Alternatively, one might think that the word 'would' is characteristic of the subjunctive mood. But not all conditionals that feature the word 'would' are intuitively Type Two. Williamson gives the following example: 'If Loretta is loyal, she would not betray a friend'. (2020, p. 170) This conditional may initially sound strange, but it is not a completely artificial construction: it may naturally be asserted by someone who endorses the more general thesis that a person is loyal only if they are disposed not to betray their friends.

Alternatively, one might argue that what philosophers really mean by 'subjunctive' in this context is just whatever grammatical mood is exemplified by conditionals like (2). But Lewis himself claims that some intuitively Type One conditionals about the future exemplify the same grammatical mood, e.g. 'If our ground troops entered Laos next year, there would be trouble'. (1973, p. 4) He concludes: 'my title 'Counterfactuals' is too narrow for my subject... but I know no better... The title 'Subjunctive Conditionals' would not have delineated my subject properly [either].' (1973, pp. 3-4)

Now let us take a look at Lewis's preferred label for Type One conditionals: 'in-

[4] See Huddleston et al (2021, p. 79).

dicative'. Like 'subjunctive', this makes reference to a grammatical mood - in this case, a mood normally used to make statements about the actual world. (1) is certainly expressed in the indicative mood, and this does indeed seem to be something that distinguishes it from (2). However, it is doubtful that the indicative mood is what distinguishes all Type One conditionals from all Type Two conditionals. As we have seen, on the assumption that conditionals like (2) exemplify the subjunctive mood, Lewis himself takes some subjunctive conditionals about the future to be intuitively Type One. Moreover, as Davis points out, there is just an 'a priori implausibility that such a grammatical difference should mark a semantic distinction of any importance'. (1979, p. 547) Grammar is hostage to linguistic tradition, and we should not expect linguistic tradition to perfectly divide our vocabulary into logical categories.

Perhaps there is some better label for Type One conditionals, but I think not. We might suggest the term 'factual', where a conditional is factual iff it is not counterfactual. This label has very occasionally been used - see e.g. Goodman (1947, p. 10). But as we saw above, Dutchman conditionals are counterfactual, despite intuitively being Type One. Hence, not all Type One conditionals are factual.

The upshot of all this is that the indicative/subjunctive and factual/counterfactual nomenclatures seem to be unsatisfactory as ways of distinguishing the two logical types of conditional. Most dualists accept this, and simply adopt some combination of these labels in the knowledge that it may be misleading to do so. For example, Bennett writes:

> Holding my nose, I adopt the labels 'subjunctive' and 'indicative'. Fortunately, their defenders never claim that 'subjunctive' helps us to understand conditionals of the type to which they apply it... The word serves only to remind them of the primacy of 'would' in (most of) the conditionals to which the label is applied... (Bennett, 2003, §5)

Of course, to monists, this is a case of the emperor's new clothes - if there *were* only one type of conditional, then this is exactly what we would expect to find! Still, the current point is merely to separate the meanings of 'indicative' and 'counterfactual' from the logical types for which they are often used as labels. This will serve us well when it comes to reconstructing and evaluating the arguments from correspondence that feature in the following sections.

5.2 An Argument from Natural Correspondence

Here is Lewis's argument for dualism as presented in his *Counterfactuals*:

> The first conditional below is probably true, but the second may very well be false...

> *If Oswald did not kill Kennedy, then someone else did.*
>
> *If Oswald had not killed Kennedy, then someone else would have.*
>
> Therefore there really are two different sorts of conditional; not a single conditional that can appear as indicative or as counterfactual depending on the speaker's opinion about the truth of the antecedent.' (D. K. Lewis, 1973, p. 3)

Let us try to reconstruct this argument so that we might more easily critique it.

As can be seen, Lewis evaluates (1) and (2) differently and then infers dualism on the basis of these different evaluations. The first evaluation ('probably true') is explicitly in terms of probability, but the second evaluation ('may very well be false') is not, and could easily be misunderstood as conveying merely that (2)'s falsehood is *possible*. However, the adverbial phrase 'very well' suggests a high *degree* of possibility, and the fact that Lewis uses the word 'but' to conjoin the two evaluations suggests that they are intended to be at odds with each other. For these reasons, one might reasonably take the first premise of Lewis's argument to be something like '(1) is probable, whereas (2) is not'.

However, given that Lewis's argument is just one of many arguments from correspondence, it would be good to identify a common structure between all these arguments so that our critique might apply to all of them. With this in mind, I suggest that we do not take the claim '(1) is probable, whereas (2) is not' as the first premise in our reconstruction. Dualism is primarily a thesis about logic, not probability, and the common thread between Lewis's argument and similar arguments for dualism is not that the relevant conditionals are probabilistically distinct; rather, it is that they are *logically* distinct. This claim may be left implicit by Lewis, but it is reasonable enough to infer it from what is made explicit.

What about the conclusion of Lewis's argument? How exactly should it be defined? First, I assume that Lewis is talking about natural language conditionals in particular (as opposed to e.g. artificial linguistic constructions that logicians have stipulated to be conditionals). Second, I also assume that the 'sorts of conditional' Lewis has in mind are *logical* sorts in particular, since he is attempting to section off a subset of conditionals to which his overarching logical thesis in *Counterfactuals* applies. Third, even though dualism is technically the view that there are exactly two logical types of natural language conditional, I assume that it would not undermine Lewis's view if it turned out that there were more than two logical types. Taking all of this into consideration, let us restate the conclusion of Lewis's argument as follows:

dualism There are (at least) two logical types of natural language conditional.

This conclusion explains the claim that (1) and (2) are logically nonequivalent, but it is not the only explanation of that claim. Most obviously, it could just be that (1) and (2) do not have the same antecedent and consequent as each other. Lewis presumably knew this, and presumably thought it was clear enough that (1) and (2) *do* have the same antecedent and consequent. Let us capture this with an implicit premise that identifies (1) and (2) as *corresponding* natural language conditionals, with 'correspondence' defined as follows:

correspondence 'If P, then Q' corresponds to 'If R, then S' iff $P = R$ and $Q = S$.

As defined, correspondence holds in virtue of identity of propositions: the antecedent and the consequent must be identical. Note, however, that it holds *between* conditional *sentences*, not conditional propositions. This allows for the possibility that the distinctness of corresponding conditionals is merely lexical. Or to put it another way: corresponding conditionals might turn out to express the very same conditional proposition.

With all this said, I propose the following reconstruction of Lewis's argument:

A Lewisian argument for dualism

P1 (1) *and* (2) *are logically nonequivalent*.

P2 (1) *and* (2) *are corresponding natural language conditionals*.

Therefore,

C *There are (at least) two logical types of natural language conditional*. (dualism)

Since P2 describes (1) and (2) as 'corresponding natural language conditionals', let us call this argument an 'argument from natural correspondence'.

This reconstruction may initially seem uncharitable, since it does not exhibit a deductively valid structure. To remedy this, one could add in an implicit conditional premise with the conjunction of P1 and P2 as antecedent and dualism as consequent. However, evaluating an argument with a conditional premise requires us to assume a set of truth conditions for the conditional. Instead, in the absence of a conditional that licenses the inference from P1 and P2 to dualism, we can simply assess the inference itself.

On that note, one thing that should be noted immediately is that dualism is not the only possible explanation of P1 even if P2 is true. According to Stalnaker's view of conditionals, for example, nonequivalent conditionals may be thought to belong to the same logical type even while having the same antecedent and consequent. As we saw in sections 2.4 and 2.5, Stalnaker (1968) argues that there is a conditional function which pairs each conditional to the most similar A-world, such that $A > C$ is true iff the A-world picked out is also a C-world. This makes conditionals variably strict, but it also makes them context-sensitive in a more complicated way thanks to the notion of *similarity* involved. According to Stalnaker:

> [T]he conditional connective is semantically unambiguous. It is obvious, however, that the context of utterance, the purpose of the assertion, and the beliefs of the speaker or his community may make a difference to the interpretation of a counterfactual. (Stalnaker, 1968, p. 109)

In other words, if corresponding conditionals are asserted in different contexts, the conditional function may pick out different A-worlds even though A is common to both conditionals. Thus, one can explain both the logical nonequivalence and the correspondence of (1) and (2) without positing two logical types of conditional.

Nonetheless, if we take the possibility of context-sensitivity off the table, then it does seem to be impossible (in some sense) for dualism to be false given the truth of P1 and P2. To give the Lewisian argument a fighting chance, let us suppose from now on that context-sensitivity is not a possibility. As we will see, even under this supposition, it is ultimately unsatisfactory.

5.3 Two Arguments from Traditional Correspondence

In two brief papers published shortly after *Counterfactuals*, E. J. Lowe (1979, 1980) argues that (2) actually corresponds to the following indicative conditional:

(3) *If Oswald* has *not killed Kennedy, then someone else* will *have*.

This conditional, he writes, might be asserted by 'someone believing that Kennedy was destined to be murdered at some prior date and suspicious, perhaps, of Oswald's intentions, but lacking direct evidence that the assassination had occurred.' (1979, p. 140) On this basis, Lowe presents two counterarguments against Lewis: one which hinges on (3) having priority over (1) as the indicative counterpart of (2), and one which alludes to an analogous argument in favour of the thesis that corresponding indicative and counterfactual conditionals are logically *equivalent*. In this section, I explain both of these counterarguments. In doing so, I identify two arguments from correspondence that are importantly different to the argument from natural correspondence reconstructed above.

To begin with, Lowe argues that (1) is material, i.e. equivalent to 'Either Oswald killed Kennedy, or someone else did', or alternatively:

(4) *Someone killed Kennedy*.

He reasons as follows. First, he notes that (4) is our ground for believing (1), and that this seems to mean that it *entails* (1). Second, he notes that (4) is *entailed* by (1), since (1) is inconsistent with (4)'s negation: 'No one killed Kennedy'. Third, he notes that if both of these entailments hold, then (1) and (4) are logically equivalent, and on this basis, he concludes that (1) is material. However, Lowe thinks that (1) is a rare example of a material conditional, and that most indicative conditionals are *non*-material. To illustrate this, he argues that (3) (an indicative conditional) is logically equivalent to (2) (a paradigmatic example of a non-material conditional). Imagining the person who might assert (3), he writes:

> [S]uch a speaker, on receiving the information that Oswald had in fact committed the crime, would instinctively amend his original assertion... by asserting (2) instead. But such an amendment would not be intended by the speaker to convey any change in his previously expressed opinion: on the contrary, it would be intended as a reaffirmation of that opinion, albeit within an altered framework of assumptions. This implies that [these conditionals] are logically equivalent and differ only in what they indicate as to the speaker's assumptions concerning the truth value of their common antecedent.' (Lowe, 1979, pp. 140-141)

On the basis of this, Lowe claims that (3) has priority over (1) as the indicative counterpart of (2).

One might reasonably wonder exactly which premise of Lewis's argument this counterargument of Lowe's is intended to undermine. Indeed, if we interpret Lowe as attacking the reconstruction of Lewis's argument presented above, then it seems he completely misses the mark: not only does he fail to undermine dualism, he *supports* it. Davis (1980) makes exactly this point:

> Lewis's conclusion was that '*there really are two different sorts of conditional*' (Lewis 1973 p. 3, quoted in Lowe 1979 p. 139)... Once it is granted, as Lowe does, that (1) and (2) are not equivalent, Lewis's conclusion follows. (Davis, 1980, p. 185)

More on this shortly.

What about Lowe's second counterargument? The idea here is that one can use (3) and (2) to construct an argument for 'the opposing thesis' (1980, p. 190), the persuasive force of which cancels out the persuasive force of Lewis's argument and thereby shows that 'nothing Lewis says gives any support to his thesis...' (Lowe, 1980, p. 190) However, this counterargument is hopeless if the opposing thesis is taken to be *monism*. Given the above definition of dualism, let us define monism as follows:

monism There is (exactly) one logical type of natural language conditional.

The persuasive force of the following argument for monism is much weaker than the persuasive force of the Lewisian argument for dualism:

A Lovian[5] argument for monism

P1 (3) *and* (2) *are logically equivalent*.

P2 (3) *and* (2) *are corresponding natural language conditionals*.

Therefore,

C *There is (exactly) one logical type of natural language conditional*. (monism)

This is a bit like inferring that all swans are white after observing two white swans. Put simply, the conclusion can very easily be false even if both premises are true.

What then has gone wrong with Lowe's counterarguments? The problem is that Lowe has identified Lewis's conclusion as 'that corresponding indicative and counterfactual conditionals require different logical treatments'. (1979, p. 140) This is a more specific claim than dualism, and it requires different premises to justify it. It is easy to see why Lowe identifies Lewis's conclusion this way, since Lewis does use the words 'indicative' and 'counterfactual' (see quote above). Importantly, however, he uses them when stating his *opponent*'s thesis that there is 'a single conditional that can appear as indicative or as counterfactual depending on the speaker's opinion about the truth of the antecedent.' (1973, p. 3) He does not use them when stating his own thesis that 'there really are two different sorts of conditional'. (1973, p. 3)

With this in mind, we should interpret Lowe as arguing against a different reconstruction of Lewis's argument, the conclusion of which may be called 'Traditional Nonequivalence Thesis', or:

TNT Corresponding indicative and counterfactual conditionals are logically nonequivalent.

To make the inference to this conclusion as strong as possible, we ought to interpret Lewis as arguing from the claim that (1) and (2) are corresponding indicative and counterfactual conditionals in particular. This gives us the following reconstruction of his argument:

A Lewisian argument for TNT

P1 (1) *and* (2) *are logically nonequivalent*.

[5] To be clear, Lowe does not defend monism, nor does he defend this argument. The eponymous adjective 'Lovian' is just a way to remind the reader that Lowe defends P1 and P2 of this argument.

P2 (1) *and* (2) *are corresponding indicative and counterfactual conditionals*.

Therefore,

C *Corresponding indicative and counterfactual conditionals are logically nonequivalent.* (TNT)

Let us call this argument an 'argument from traditional correspondence'. This reconstruction may initially seem like a more charitable one than the argument from natural correspondence above, since it *sounds* deductively valid. However, as we will see in the next section, it is neither deductively valid nor charitable.

If we interpret Lowe as attacking this argument from traditional correspondence, we can begin to make sense of his counterarguments. The first counterargument makes sense as an attempt to undermine P2 of this reconstruction, the implication being that (1) and (2) are *not* corresponding indicative and counterfactual conditionals because (3) 'has prior claim to be regarded as the indicative counterpart of (2)'. (1979, p. 140) And the second counterargument makes sense as alluding to the possibility of using (3) and (2) to construct an argument for what we might call 'Traditional Equivalence Thesis', or:

TET Corresponding indicative and counterfactual conditionals are logically equivalent.

The argument alluded to is as follows:

Lovian argument for TET

P1 (3) *and* (2) *are logically* equivalent.

P2 (3) *and* (2) *are corresponding indicative and counterfactual conditionals*.

Therefore,

C *Corresponding indicative and counterfactual conditionals are logically equivalent*. (TET)

Like the argument for TNT, let us call this argument an 'argument from traditional correspondence'.

Even when interpreted as attacks on the Lewisian argument for TNT, both of Lowe's counterarguments have serious problems. Consider again his first counterargument. If we grant that (1) is material and that (3) is logically equivalent to (2), nonetheless there is no reason to think that (3) 'has prior claim to be regarded as *the* indicative counterpart of (2).' (1979, p. 140, my italics) To describe (3) this way presupposes that there can be only one indicative counterpart of (2), yet there is nothing in the concept of correspondence that prohibits multiple counterparts. In other words, there is nothing to stop us saying that (1) and (3) are *both* indicative counterparts of (2). One might think that there can be only one indicative counterpart of (2) if one thinks that only one conditional proposition per logical type can be constructed out of *A* and *C*. But as we have seen, 'indicative' does not mean Type One, nor does 'counterfactual' mean Type Two, and dualists tend to be aware of this. Indeed, if these labels did refer to different logical types, there would be much less room for debate regarding TNT and TET.

It may seem that this objection to Lowe's first counterargument is made possible only because we have defined correspondence as holding between conditional sentences - if one thinks of correspondence as holding between different conditional *propositions*, then each member of a set of corresponding conditionals must belong to a different logical type (since we are supposing that Stalnakerian context-sensitivity is impossible). Thinking of correspondence this way limits us to a maximum of two corresponding conditionals *if* the possibility of three or more logical types of conditional is off the table. Besides, Lowe himself would most likely have approved of my definition of correspondence. He writes:

> The sort of 'correspondence' presupposed by Lewis's thesis must obviously be that which exists between a non-material indicative conditional and a counterfactual conditional having *the same antecedent and consequent*. Now in ordinary language no such correspondence strictly speaking ever exists, at least at a purely lexical level. Nonetheless, one may hope to discover pairs of such conditionals whose antecedents and consequents 'express the same propositions' (however one may interpret that controversial phase [sic]). (1980, p. 189)

Here, Lowe clearly expresses a preference for thinking of corresponding conditionals and their constituent parts as lexical items first and foremost. He moves away from thinking of antecedents and consequents as lexical items only because he recognises that we cannot think of them this way if we hope to find any pairs of corresponding conditionals.

Now consider again Lowe's second counterargument, according to which the persuasive force of the Lovian argument for TET cancels out the persuasive force of the Lewisian argument for TNT. I think this counterargument is more promising, but as I will explain, Lowe does not quite do it justice. In order to know precisely how to make sense of this counterargument, we need to know whether Lowe considers TNT to be the sort of thing that can be true despite the existence of counterexamples. Statements like TNT are known as generics, and the hallmark of a generic is that it generalises without specifying a quantity or proportion - TNT does not say, for example, that *all* corresponding indicative and counterfactual conditionals are logically nonequivalent. In general, generics seem to be consistent with the existence of counterexamples. For instance, 'Cats have tails' seems to be true despite the existence of Manx cats. In the same vein, Lowe writes that the equivalence of (2) and (3) 'does not prove that [TNT] is false'. (Lowe, 1980, p. 190) Hence, Lowe does not think TNT can be refuted by a single counterexample.

On the other hand, Lowe does not seem to see TET the same way. Initially, he writes that Lewis's conclusion 'fails to follow because one can find *another* indicative conditional which is both verbally close to (2) and logically equivalent to it.' (1979, p. 139) This suggests that Lewis's conclusion *would* follow otherwise, which suggests that TNT can be *proven* by a single example, which in turn suggests then TET can be *refuted* by a single counterexample, since TNT and TET are incompatible. Let us put all this information into a table so that it may more easily be remembered:

Lowe's interpretation	TNT	TET
Can it be proven by one example?	Yes	No
Can it be refuted by one counterexample?	No	Yes

The trouble is, if we interpret these theses this way, then the persuasive force of the argument in favour of TET does not cancel out the persuasive force of the argument in favour of TNT. In order for that to happen, the inference in the argument for TET would have to be at least as strong as the inference in the argument for TNT. But on this interpretation, the inference in the former argument is much weaker than the inference in the latter: TNT *can* be proven by one example, but TET *cannot*. In the next section, I suggest some alternative interpretations of the two theses. On each of these alternative interpretations, Lowe's second counterargument can be salvaged.

5.4 Evaluating Arguments from Traditional Correspondence

With a view to minimising confusion going forward, let us summarise the lessons that can be drawn from evaluating Lowe's counterarguments. Firstly, we can distinguish between different relations of correspondence depending on the way in which the relata are characterised: some correspondence relations hold between natural language conditionals of any kind, whereas others hold between more specific kinds such as indicative and counterfactual conditionals. One might object that having relata of specific kinds does not mean the relation of correspondence itself is of a different kind. Fair enough - I speak otherwise only because it helps when categorising arguments from correspondence. Secondly, the conclusions of these arguments from correspondence vary depending on the type of correspondence involved: arguments from natural correspondence do not support conclusions about specific types of natural language conditional, whereas arguments from traditional correspondence do. Lastly, the persuasive force of the latter depends in part on how we interpret their conclusions: if we interpret them as being such that a single example can prove them, then the support provided by the premises is strong; otherwise, it is weak.

The exact truth conditions of generics is something about which there is much debate. However, I think Lowe's interpretation of TNT and TET is especially strange, because it treats the two theses unequally even though both are about the same subject matter: the matter of whether corresponding indicative and counterfactual conditionals are logically equivalent. If TNT and TET were dualism and monism respectively, then Lowe's interpretation would make more sense. But TNT and TET are not dualism and monism respectively.

In this section, I look at the four possible interpretations of TNT and TET according to which both theses are equally easy/difficult to prove/refute. On all four interpretations, Lowe's second counterargument can be salvaged: if TNT and TET are equally easy/difficult to prove, then the inferences in their respective arguments must be equally strong. Moreover, as we will see, the most plausible of these four interpretations leads us to an objection against *all* arguments from traditional correspondence.

Generally speaking, when considering whether two propositions are logically equivalent, there should be an assumption of nonequivalence until proven otherwise, because nonequivalence is the norm. However, this does not mean that the matter of *proving* nonequivalence should be any easier than proving equivalence. Moreover, the assumption of nonequivalence goes out the window when we are considering pairs of corresponding conditionals in particular - in their case, if anything, there ought to be an assumption of *equivalence* until proven otherwise, because dualism is a more complicated view than monism. Since Lowe's interpretation treats TNT as both easier to prove *and* harder to refute than TET, I think we should reject it. Instead, let us consider the four possible interpretations that treat TNT and TET *equally*:

interpretation 1	TNT	TET
Can it be proven by one example?	Yes	Yes
Can it be refuted by one counterexample?	No	No

interpretation 2	TNT	TET
Can it be proven by one example?	No	No
Can it be refuted by one counterexample?	Yes	Yes

interpretation 3	TNT	TET
Can it be proven by one example?	Yes	Yes
Can it be refuted by one counterexample?	Yes	Yes

interpretation 4	TNT	TET
Can it be proven by one example?	No	No
Can it be refuted by one counterexample?	No	No

Interpretation 1 would be the right interpretation of TNT and TET if both theses began '*Some* corresponding indicative and counterfactual conditionals...' Compare, for example, 'Some cats have tails' and 'Some cats do not have tails', each of which can be proven, but not refuted, by a single cat (although not the *same* cat). However, interpretation 1 can be rejected immediately. Without the word 'Some', the two theses are incompatible: they cannot both be true. Compare 'Cats have tails' and 'Cats do not have tails'. When two theses are incompatible, proving one means refuting the other; hence, if one of these two theses can be proven by a single pair of conditionals, then the other can be refuted by that same pair of conditionals. This rules out interpretation 1, since it has a 'Yes' on the top row and a 'No' on the bottom row.

None of the other interpretations can be ruled out on this basis, because none of them have implications for the possibility of refuting one thesis by proving the other. However, interpretation 2 does have implications for the possibility of *proving* one thesis by *refuting* the other. In particular, it implies that the refutation of one thesis is *insufficient* to prove the other, since it has a 'Yes' on the bottom row but a 'No' on the top row. Under this interpretation, it is not that both theses can be true at the same time; rather, it is that both theses can be *false* at the same time. It is of course perfectly possible that neither thesis is true. Compare, for example, 'Indicative conditionals are true' and 'Indicative conditionals are not true', both of which are false. Here are some other examples: 'Humans are female' and 'Humans are not female'; 'Planets are gaseous' and 'Planets are not gaseous'; 'Cheese is smoked' and 'Cheese is not smoked'...

So far, then, interpretation 2 seems to be in good standing. Yet there is reason to be suspicious of it. In particular, it makes TNT and TET very easy to disprove: just a single counterexample will do. This assumes that the matter of whether corresponding indicative and counterfactual conditionals are logically equivalent is importantly different to e.g. the matter of whether cats have tails: the generic 'Cats have tails' is true in a world where some cats do not have tails, whereas under interpretation 2, neither TNT nor TET can be true in a world where only some corresponding indicative and counterfactual conditionals are logically nonequivalent. Some might struggle to imagine a world in which only some corresponding indicative and counterfactual conditionals are logically nonequivalent. Yet, on any sensible version of dualism, corresponding indicative and counterfactual conditionals whose consequents are the same as their antecedents are logically equivalent. For example, on Lewis's view, $P \rightarrow P \equiv P \mathrel{\Box\!\!\rightarrow} P$. For this reason, we should not adopt an interpretation according to which TNT and TET can each be refuted by a single counterexample. That means we should not adopt interpretation 2.

In fact, it also means that we should not adopt interpretation 3, since this interpretation also has a 'Yes' on the bottom row. We are left, then, with interpretation 4. This interpretation does not imply that TNT and TET exhaust the logical space, nor does it imply that the subject matter is importantly different to other subject matters about which we can state generics. And if we do adopt interpretation 4, then there is a much more obvious objection to be made against the argument for TNT. (Indeed, the fact that Lowe does not make it is a fairly strong reason to think that he did not have this interpretation in mind.) That objection is simple: the inference in the argument for TNT is just *extremely weak*. The argument attempts to justify a generic on the basis of a single example. It is no better than an argument from the existence of a single Manx cat to the generic 'Cats do not have tails'.

This gives us a compelling objection against *all* arguments from traditional correspondence: whether they are in favour of logical equivalence or nonequivalence, the inferences involved in these arguments are extremely weak. The fact that two things of certain types bear/do not bear a certain relation to each other simply does not mean that things of those types *in general* bear/do not bear that relation to each other. Whether one argues for TNT or TET, a single example simply cannot prove the conclusion.

Interestingly, this counterargument works also under interpretation 2, since that interpretation has a 'No' on the top row. Of the three interpretations that do not lead to the absurdity that TNT and TET are compatible, only one has a 'Yes' on the top row: interpretation 3. If one adopts interpretation 3, then this counterargument does not work. However, Lowe's first counterargument *does* work. Interpretation 3 does not immediately lead to absurdity, because it is consistent with the idea that proving one thesis means refuting the other, but it does lead to absurdity *if* we suppose that all the premises of the arguments for TNT and TET are true. Under interpretation 3, both arguments suffice to prove their conclusion; additionally, both arguments suffice to refute the opposing thesis. So, on interpretation 3, it cannot be that both arguments have true premises. That puts the proponent of interpretation 3 in a position where they have to choose which of the two arguments (if either) to endorse. If they endorse only one of the two arguments, then it seems to me that they ought to endorse the argument for TET, because (as Lowe contends) the case for treating (3) as an indicative counterpart of (2) is much stronger than the case for treating (1) as such. Lowe writes:

> It may be noticed that even at the lexical level the resemblance between [(3)] and (2) is considerably closer than that between (1) and (2), since the former pair differ merely in that 'has' is exchanged for 'had' and 'will' for 'would'. More importantly, at the syntactical level, the tense structure of [(3)] parallels that of (2) in a way which that of (1) does not. (Lowe, 1979, p. 139)

The proponent of interpretation 3 might argue that P1 of the argument for TNT is more plausible than P1 of the argument for TET. But it is hard to see why one would argue this. And as above, when dealing with pairs of corresponding conditionals, there ought to be an assumption of logical *equivalence*, not nonequivalence. So one ought to start from the assumption that the conjunction of P1 and P2 of the argument for TNT is less plausible than the conjunction of P1 and P2 of the argument for TET.

It may be noted that on interpretation 3, both of Lowe's counterarguments can be salvaged. Perhaps this was the interpretation he had in mind all along, despite saying things that are inconsistent with it. If so, this section can be seen as a clarification of Lowe's counterarguments. It should not, however, be seen as a *defence* of those counterarguments. As above, I think we should reject interpretation 3 in favour of interpretation 4. On this interpretation, Lowe's first counterargument fails, and his second counterargument is made redundant by the fact that all arguments from traditional correspondence involve extremely weak inferences, including the Lovian argument for TET.

5.5 Evaluating Arguments from Natural Correspondence

In section 5.3 above, I stated a Lovian argument from natural correspondence in favour of monism:

> **A Lovian argument for monism**
>
> P1 (3) *and* (2) *are logically equivalent*.
>
> P2 (3) *and* (2) *are corresponding natural language conditionals*.
>
> Therefore,
>
> C *There is (exactly) one logical type of natural language conditional*. (monism)

As mentioned at the time, the persuasive force of this argument is much weaker than the analogous Lewisian argument for dualism - that is why Lowe's second counterargument fails as an attack on the Lewisian argument for dualism. Indeed, any argument from natural correspondence in favour of monism will involve an extremely weak inductive inference. For this reason, arguments from natural correspondence in favour of monism can immediately be dismissed. Instead, let us focus our attention on arguments from natural correspondence in favour of dualism.

My objection to arguments from natural correspondence in favour of dualism is simple: they do not appeal to beliefs that are common to the arguer and their opponent. This does not mean that such arguments are unsound, but it does mean that they are no way to make progress on the topic. Like arguments from traditional correspondence, arguments from natural correspondence have a simple, syllogistic structure: two premises and a conclusion. Such arguments can easily be flipped on their head. Consider again the Lewisian argument for dualism:

> **A Lewisian argument for dualism**
>
> P1 (1) *and* (2) *are logically nonequivalent*.
>
> P2 (1) *and* (2) *are corresponding natural language conditionals*.
>
> Therefore,
>
> C *There are (at least) two logical types of natural language conditional*. (dualism)

With Stalnakerian context-sensitivity off the table, we are assuming that the conjunction of P1, P2 and ¬C is impossible. If so, one can just as easily reason from ¬C and P2 to ¬P1, or from ¬C and P1 to ¬P2. This much is not unusual: one man's *modus ponens* is another man's *modus tollens*, as they say. But monists are bound already to deny one of these premises, even if the need to do so has not occurred to them. Once Stalnakerian context-sensitivity is off the table, it is just an obvious and immediate consequence of monism that logically nonequivalent conditionals cannot have the same antecedent and consequent as each other. The upshot of this is that arguments from natural correspondence in favour of dualism are unpersuasive to monists, because they do not appeal to beliefs that are common ground. In other words, they preach to the choir.

One might object that this is just a consequence of the way that I have reconstructed Lewis's argument. After all, Lewis makes reference to the apparent difference in probability between (1) and (2), and this may be construed as an attempt to appeal to common ground. Perhaps it is an attempt to make such an appeal, but if so, it is a failed attempt, because whether P1 is justified by the apparent difference in probability depends on what type of probability function we are talking about. If the probability function is *coherent* - that is, if it follows the rules of probability logic - then it will conform to the rule that equivalent propositions are equiprobable. In that case, a difference in probability justifies P1. Yet we are given no reason to think that the probability function *is* coherent. In a context like this, one can only be confident that one assigns different *subjective* probabilities to (1) and (2), and subjective probabilities are often incoherent. If one assumes that one's subjective probability function conforms to the rule that equivalent propositions are equiprobable, then one must also assume that one can reliably identify logically equivalent propositions when one sees them. But that is precisely the point of contention. Some degree of doubt regarding the probabilities of (1) and (2) is appropriate, and since one's degree of doubt depends on one's acceptance of P1, Lewis will not persuade those who disagree with P1 by appealing to an apparent difference in probability between (1) and (2).

What about the alternative claim that it is possible that (1) and (2) do not have the same truth values? Could this be considered common ground? Unfortunately, the same sort of criticism applies also to this claim: whether it justifies P1 depends on whether the modality in question is subjective. It may be that there is a conception of the epistemic modality according to which it is possible that (1) and (2) do not have the same truth values. For instance, it may be that their having different truth values is logically consistent with what one *believes*. But analogously, for some people, it is an epistemic possibility that '*A and B*' and '*A but B*' do not have the same truth values, yet that does not mean that the two propositions are logically nonequivalent. What matters is whether a difference in truth value is a *logical* possibility, and we are given no reason to think that it is.

None of this is intended to dissuade dualists of their belief in P1 of the Lewisian argument for dualism. Dualism is a coherent view, and the premises of the Lewisian argument for dualism can be supported from within that view. But if context-sensitivity is off the table, then they cannot both be supported from outside that view. Importantly, all arguments from natural correspondence in favour of dualism are susceptible to this same critique, since the only difference between such arguments is the pair of corresponding conditionals to which they appeal. It would be a mistake, therefore, to think that some new example will persuade those who have not yet been persuaded - if one is convinced of monism, then one is bound already to deny that logically nonequivalent conditionals can have the same antecedent and consequent as each other.

It may help to draw an analogy with an argument from the natural correspondence of *conjunctions*. Consider the following pair of conjunctions:

(5) *John is 12 years old and he likes opera.*

(6) *John is 12 years old but he likes opera.*

Now consider the following argument:

An argument for dualism regarding conjunctions

P1 (5) *and* (6) *are logically nonequivalent.*

P2 (5) *and* (6) *are corresponding natural language conjunctions.*

Therefore,

C *There are (at least) two logical types of natural language conjunction.*

Clearly, this argument is not going to persuade those who already disagree with the conclusion, because rejecting at least one of the premises is an obvious and immediate consequence of rejecting the conclusion. If one believes that there is just one logical type of conjunction, then one is bound already to deny that logically nonequivalent conjunctions can have the same conjuncts.

Suppose I am right about arguments from natural correspondence in favour of dualism. In that case, we require an explanation of the fact that such arguments sometimes seem to succeed in persuading people of dualism. The explanation, I believe, is just that some people do not realise their own inclination towards dualism until coming across such arguments. People can hold beliefs without realising it, and when those with inclinations towards dualism first come across an argument from natural correspondence in favour of dualism, it helps them to identify their beliefs. If I am right, we ought not to focus on arguments from correspondence in order to settle the debate between monism and dualism. Rather, we ought to focus on comparing the wider views that affirm or deny the premises of these arguments.

5.6 The Possibility of Logical Equivalence

Even if I have succeeded in showing that the dualist ought to expect the monist to reject the conjunction of premises in the Lewisian argument for dualism, one might still reasonably wonder whether it is defensible to reject the first premise in particular. Can we really make sense of the idea that (1) and (2) are logically equivalent? For most of us, when we first encounter (1) and (2), the former strikes us as much more sensible than the latter - that much I am willing to concede. In the next section, I argue that this can be explained by perceived differences relevant to the matter of assertability. However, in this section, I try to show that we can indeed make sense of the claim that (1) and (2) are logically equivalent.

One way to make sense of this claim is to take Lowe's argument for the claim that (2) and (3) are logically equivalent and extend it to form an argument according to which (1) and (3) are *also* logically equivalent. Suppose Oswald planned Kennedy's assassination with three backup assassins: Jones, Smith, and Peterson. Suppose also that their plan was to wait at different points on the motorcade route, and to behave as follows: if Oswald fails to kill Kennedy by time t_1, then Jones will attempt it between t_1 and t_2, and if Jones fails to kill Kennedy by time t_2, then Smith will attempt it between t_2 and t_3, and so on. Lastly, suppose that Peterson tells her partner, Mr. Peterson, about the plan in advance. He approves, and is confident that the plan will succeed, so decides to wait at home while the plan is being carried out. At t_4, without any information as to whether the plan has been successful, Mr. Peterson says to himself:

(3) *If Oswald has not killed Kennedy, then someone else will have.*

Then, at t_5, news of Kennedy's death reaches Mr. Peterson through his television set. He says to himself:

(1) *If Oswald did not kill Kennedy, then someone else did.*

Just as in Lowe's thought experiment, 'such an amendment would not be intended by the speaker to convey any change in his previously expressed opinion: on the contrary, it would be intended as a reaffirmation of that opinion, albeit within an altered framework of assumptions,' (Lowe, 1979, p. 140) the altered framework being one that now includes testimony of Kennedy's death. If this style of argument is persuasive, then it gives us reason to think that (3) is logically equivalent to (1) *as well as* (2), and that implies that (1) and (2) are logically equivalent. But even if this style of argument is unpersuasive, it should at least help the reader to make sense of the idea that (1) and (2) are logically equivalent.

If (1) and (2) are logically equivalent, then they are surely both *non*-material. But of course, if we are to accept that (1) is non-material, we must resist Lowe's argument in favour of its logical equivalence with (4), i.e. 'Someone killed Kennedy'. As we saw above, Lowe thinks that (4) entails (1) because it is our *ground* for believing (1). However, as Davis points out, 'not all grounds are deductive'. (1980, p. 184) In defence of this claim, Davis presents a thought experiment in which (1) seems to be false even though (4) may be true:

> Kennedy died because of a blow to the head. Either Kennedy accidentally tripped and fell on a rock or Oswald clobbered Kennedy with the rock and made it look like he tripped. No one other than Oswald was within a fifty mile radius of Kennedy at the time of his death. (Davis, 1980, p. 184)

There is a natural interpretation of (1) such that it is false in this thought experiment, even if (4) is true; that is, even if someone killed Kennedy. On this interpretation of (1), it is therefore non-material.[6]

While Davis agrees that (1) is non-material, he nonetheless does not agree that (1) and (2) are logically equivalent. Instead, he takes it for granted that (1) is true in the actual world and then tailors his non-material interpretation of (1) to accommodate this truth value. (1979, p. 549) However, I think it is possible to hear (1) as false in the actual world. Non-material interpretations of conditionals invariably imply that a conditional is true *only if* its consequent is true in the closest world in which its antecedent is true, and it is perfectly reasonable to think that the closest world in which Oswald does not kill Kennedy is not a world in which someone else does. Davis tries to persuade us against this intuition, writing:

> The assassination of Kennedy is obviously one of the most important events of recent American and world history. The fact that Oswald in particular was the assassin, rather than some other very unimportant person, is in contrast

[6] Unsurprisingly, Lowe disagrees that this interpretation is available, but that is because he is unwilling to give up the supposition that (1) is material. He writes: '(1) *is* true, simply because it *is* a material conditional and has a false antecedent and consequent.' (1980, p. 187)

> of minor significance. Consequently, a world in which someone else killed Kennedy is considerably more similar to the actual world than a world in which Kennedy was not killed. (Davis, 1979, p. 549)

But, as Davis himself notes, Lewis asserts the exact opposite intuition:

> [W]orlds where Oswald did kill Kennedy... are worlds to which worlds with no killing are closer than worlds with a different killer.' (D. K. Lewis, 1973, p. 71)

I do not expect this to persuade the reader that (1) and (2) are logically equivalent. However, it should suffice to show that we can at least make sense of the idea.

The idea that (1) and (2) are logically equivalent is made more palatable by the fact that we can identify something else about them that explains our different reactions to them: in the next section, I suggest that, when reading the two conditionals, we imagine different possibilities to be live options, and this in turn gives rise to a difference of assertability.

5.7 A Difference of Assertability

In section 4.4, I defended a norm of assertion for indicative conditionals (called KNAC) according to which the assertion of an indicative conditional is proper only if one knows that its corresponding material conditional is true at all worlds of interest to the participants in the relevant conversation. Let us suppose that this norm governs the assertion of (2) as well as (1); indeed, let us suppose that it governs the assertion of natural language conditionals in general.

Unlike Type One conditionals, people tend to assert only those Type Two conditionals that they think may be used as inference-tickets to truth at *other* possible worlds. However, there is widespread scepticism regarding our ability to gain knowledge about specific counterfactual worlds, and rightly so. Not only are counterfactual worlds causally isolated from us (supposing they exist), our conceptions of them are usually radically incomplete. This in turn hinders our ability to distinguish them from each other. In other words, when conceiving of counterfactual worlds in conversation with each other, we can rarely be confident that we are conceiving of the same things. As Quine famously puts it:

> What traits of the real world to suppose preserved in the feigned world of the contrary-to-fact antecedent can be guessed only from a sympathetic sense of the fabulist's likely purpose in spinning his fable. (Quine, 1960, p. 222)

I think this epistemic difference points to a difference of assertability between (1) and (2), as I will explain.

Neither (1) nor (2) is asserted by Lewis; rather, Lewis asserts something *about* (1) and (2). Nonetheless, when we read (1) and (2), we imagine them being asserted by someone. Let us begin with (1). I think that when we read (1), we imagine that it is asserted in a conversation where the participants know that *someone* killed Kennedy. In this imaginary situation, the set of live possibilities (i.e. the set of worlds logically consistent with the participants' common knowledge) therefore does not include any world in which no one killed Kennedy. In other words, 'Either Oswald killed Kennedy or someone else did' is a *live necessity*. Beyond this, we might fill out the imagined assertion of (1) in different ways. For instance, we might imagine that all participants in the conversation know that *Oswald* in particular killed Kennedy. In that case, none of the participants can use (1) as an inference-ticket to infer anything about the actual world: they cannot perform *modus ponens* given a false antecedent, and *modus tollens* will only tell them something that they already know. This illustrates the fact that an indicative conditional is only useful as an inference-ticket to truth at the actual world for someone who does not know the antecedent's negation. As Adams writes: 'indicative conditional statements are seldom made in the *knowledge* that their antecedents are false. The reason is clear: such an assertion would be misleading.' (1965, p. 176)

On the other hand, we might imagine that some participants in the conversation do not know that Oswald in particular killed Kennedy. In that case, (1) may be used as an inference-ticket *if* the participants in question come to know either that Oswald did not kill Kennedy or that no one else did. Of course, this knowledge is hard to come by, so it seems more likely that (1) would be asserted simply as a way of trying to narrow down the set of worlds of interest.[7] Nonetheless, it is important to note that, in this imaginary situation, (1) *can* be used as an inference-ticket to truth at the actual world. Hence, the participants in the conversation need not direct their attention to counterfactual worlds. For this reason, when reading (1), I think we imagine that the actual world is the only world of interest. That makes it very easy for the asserter to assert (1) without flouting KNAC: all they need to know is that 'Either Oswald killed Kennedy or someone else did' is true at the actual world.

Now let us consider (2). When we read (2), we imagine that it is asserted in a conversation where the set of live possibilities is relevantly different. The assertion of

[7] If one asserts a conditional even though one does not know that its corresponding material conditional is true at all worlds of interest, nonetheless the participants in the conversation may restrict their interest to a set of worlds small enough to make one's assertion proper. Plausibly, this is a mechanism exploited also by so-called 'biscuit conditionals', e.g. 'There are biscuits on the counter if you want some'. Asserting this can be seen as a way of trying to restrict interest to just those worlds in which there are biscuits on the counter. More generally, the assertion of a biscuit conditional $A \to C$ can be seen as a way of trying to narrow down the set of worlds of interest to just those in which C is true.

(2) plausibly conveys a belief in the antecedent's negation, i.e. a belief that Oswald did kill Kennedy. More than this, I think it conveys a belief that *everyone* in the conversation *knows* that Oswald killed Kennedy. For this reason, we imagine 'Oswald killed Kennedy' to be a live necessity. But in this imaginary situation, (2) is of no use as an inference-ticket to truth at the actual world: as above, the participants cannot perform *modus ponens* given a false antecedent, and *modus tollens* will only tell them something that they already know. Instead, if (2) is to be used as an inference-ticket, it must be used as an inference-ticket to truth at some other possible world(s). For this reason, when reading (2), I think we imagine that the actual world is *not* the only world of interest. That makes it relatively hard for the asserter to assert (2) without flouting KNAC, because they need to know that 'Either Oswald killed Kennedy or someone else did' is true at the relevant counterfactual world(s). Moreover, upon reading (2), many of us imagine that this statement is *false* at the relevant counterfactual world(s), and that makes it *impossible* for the asserter to assert (2) without flouting KNAC.

In sum, the proper assertion of (1) is easier than the proper assertion of (2) in the situations in which we imagine them being asserted. Importantly, this is not just some quirk of our imaginative tendencies. Rather, this follows from the fact that (2) conveys something that (1) does not: namely, the belief that everyone in the conversation knows that the antecedent is false. I do not think this belief is conveyed by *all* conditionals that the dualist categorises as counterfactual (see e.g. Watson's conditional above), but I do think it is conveyed by (2).

5.8 Conclusion

In this chapter, I have argued that we should look beyond arguments from correspondence when answering the question 'How many different logical types of natural language conditional are there?' I began by supposing the truth of dualism and evaluating different labels for the two logical types. I then reconstructed Lewis's argument as an argument from natural correspondence and explained Lowe's counterargument to it. Doing so involved re-interpreting Lewis as presenting an argument from traditional correspondence and laying out an analogous argument for the opposing thesis. I hope to have persuaded the reader that the inferences in arguments from traditional correspondence are extremely weak, whereas arguments from natural correspondence either involve weak inferences or preach to the choir. I hope also to have shown the reader that we can make sense of the idea that (1) and (2) are logically equivalent. This idea is made more palatable by the fact that we can still explain our different reactions to (1) and (2) by appeal to a difference of assertability.

It was said in section 2.5 that we might reasonably hope to find a Y-shaped analysis of conditionals: one which identifies both the differences *and* commonalities of conditionals like (1) and (2). The previous section has shown that the metaphysically strict analysis has "branches". In the next chapter, I aim to shed light on its "trunk".

Chapter 6

A Strict Analysis of 'Would'-Conditionals

Scepticism regarding counterfactual conditionals has been promoted by many influential philosophers. For example, Quine writes: '[T]he subjunctive conditional is an idiom for which we cannot hope to find a satisfactory general substitute in realistic terms...' (1960, p. 222) If Quine is right, it is a shame, because such conditionals are very useful. For example, they can be used to understand dispositional properties. As Quine himself explains, '[t]o say that *a* is *fragile*... is to say that if *a* were struck smartly... *a* would break'. (1960, pp. 222-223) They can also be used to understand causal dependence (see e.g. Lewis (1973)) and causal explanation (see e.g. Woodward (2003)). The list goes on. As Hájek writes, they 'figure in influential analyses of... perception, knowledge, personal identity, laws of nature, rational decision, confirmation... free action... and so on.' (2014, p. 5) It would be nice, then, to have a view of them according to which they are philosophically respectable.

To make a strong defence of a theory of indicative conditionals, one ought to show that it fits with a theory of counterfactual conditionals, or at least with a theory of *some* of the conditionals commonly labelled as 'counterfactual'. In this chapter, I argue that 'would'-conditionals (i.e. conditionals that use the word 'would' in their consequent) may reasonably be seen as having exactly the same truth conditions as indicative conditionals: they are true iff their corresponding material conditionals are metaphysically necessary. My aim is not to persuade the reader beyond doubt that these are the truth conditions of 'would'-conditionals; rather, I have the more modest aim of demonstrating that the metaphysically strict analysis may be extended without absurdity to cover a large proportion of counterfactual conditionals, thereby giving us what may reasonably be regarded as a unified analysis of natural language conditionals in general. To the extent that this unified analysis is plausible, the metaphysically strict analysis of indicative conditionals is strengthened.

To begin, I argue alongside Williamson that 'would' may be understood as a necessity operator capable of featuring in unconditional propositions. I explain Williamson's view that 'would' is a *circumstantial* necessity operator that ranges over contextually relevant subsets of metaphysically possible worlds. In section 6.2, I go over some important concepts in the possible worlds calculus that enable us to more carefully investigate the consequences of combining a Williamsonian view of 'would' with a metaphysically strict analysis of 'would'-conditionals. The resultant interpretation of 'would'-conditionals is not totally implausible, but it attributes to them a logical structure that is quite unwieldy. I suggest an independently motivated simplification that involves treating 'would' as a *metaphysical* operator taking both the antecedent *and* consequent of a 'would'-conditional within its scope. When seen this way, the 'would' operator makes no difference to the conditional's truth conditions. Of course, one consequence of this interpretation is that a great many 'would'-conditionals are false, but as we have seen (in chapter 4), that does not mean that they are unassertable. Before concluding, I show that this view of 'would'-conditionals coheres also with a plausible view of 'might'-conditionals, such as 'If Oswald had not killed Kennedy, then someone else might have'. According to the view defended, 'might'-conditionals are just *negated* metaphysically strict conditionals.

6.1 The 'Would' Operator

Consider (again) the pair of conditionals found at the beginning of Lewis's *Counterfactuals* (1973, p. 3):

(1) *If Oswald did not kill Kennedy, then someone else did*.

(2) *If Oswald had not killed Kennedy, then someone else would have*.

With these conditionals in mind, Williamson writes:

> Methodologically, it is odd to explain the truth-conditional difference between two sentences by postulating something like ambiguity in a part where they look exactly the same, while marginalizing the visible differences elsewhere. The natural default hypothesis is that 'if' means exactly the same in (1) and (2)... (Williamson, 2020, p. 167)

Ultimately, I do not agree that there is a truth-conditional difference to be explained, but I do agree that we ought to consider the possibility that the two conditionals have different constituent propositions. That is what we will do in this section.

There is a traditional thesis (defended e.g. by Kasper (1992)) that when 'would' features in an unconditional statement, it is really the consequent of an implicit conditional. In Kasper's words: 'The traditional view on simple sentences in subjunctive mood regards them as a kind of counterfactual conditional with a missing antecedent.' (1992, p. 307) Suppose, for example, that I am considering how to get down from a roof without a ladder. While staring over the edge, I might say to myself, 'I would break my leg'. According to the traditional thesis, this is really the consequent of an implicit conditional - something like:

(3) *If I were to jump, I would break my leg.*

Williamson rejects this traditional thesis, arguing instead that unconditional 'would'-statements make good sense on their own. He gives the following example (2020, p. 168):

(4) *Loretta would not betray a friend.*

Williamson's view is that (4) can be analysed as a modal proposition, where 'would' is a 'restricted *necessity* operator' (2020, p. 169), and the proposition expressed by 'Loretta does not betray a friend' falls within its scope.

In order to combine this view of 'would' with a metaphysically strict analysis of 'would'-conditionals, we need to know in what sense Williamson considers the necessity operator to be 'restricted'. He writes:

> What sort of modality does 'would' express? It is not epistemic; (4) may in fact be true even though no relevant body of knowledge entails that Loretta does not betray a friend... Nor is the modality deontic... it is not part even of what (4) says that Loretta *ought* not to betray a friend. Rather, 'would'... is *circumstantial*; it ranges over objective possibilities, determined by how things are. That is not to say that [unconditional 'would'-statements] attribute metaphysical necessity to the complement of 'would'... Even though it is metaphysically possible for Loretta to betray a friend, (4) may still be true. The implicit universal quantifier in (4) is implicitly restricted to contextually relevant worlds, in some sense. (Williamson, 2020, pp. 169-170)

In summary, Williamson thinks the 'would' operator is a universal quantifier ranging over subsets of metaphysically possible worlds, and that these subsets are determined by context. Let us say that he takes 'would' to be a *circumstantial* necessity operator, defining the circumstantial modality as follows:

circumstantial modality *P* is circumstantially possible iff it is contextually relevant.

I do not agree with everything Williamson has to say about 'would', but I do think he is right that we can make sense of it independently of conditionals, as demonstrated by unconditional 'would'-statements like (4). More specifically, I agree that 'would' is a necessity operator. However, combining this view of 'would' with a metaphysically strict analysis of 'would'-conditionals is complicated: the result is an interpretation of 'would'-conditionals as having a necessity operator within the scope of another necessity operator. In order to evaluate this analysis, we need to know how to interpret propositions that have modal operators within the scope of other modal operators. This is sometimes known as 'iterated' modality (see e.g. Lewis (1986a, p. 18)), but I prefer the term 'nested', because it makes clear that one modal operator is within the scope of the other. Let us use the term 'simple nested modal proposition' to describe a proposition involving nested modality where the operators express the same modality, and let us use the term 'complex nested modal proposition' to describe a proposition involving nested modality where the operators express *different* modalities. In section 6.3, I look at some examples of the latter, but in the next section, I look exclusively at examples of the former.

6.2 Simple Nested Modality

Consider the statement 'It is metaphysically possible that P is metaphysically possible'. This statement is grammatical, but its assertion seems infelicitous - intuitively, there is one modal operator too many. The same can be said for the assertion of 'It is metaphysically necessary that P is metaphysically necessary'. In what follows, I use the possible worlds calculus to interpret these propositions in a way that explains why their assertions seem infelicitous.

The possible worlds calculus gives us a systematic way of interpreting nested modal propositions. Recall that $\Diamond P$ is true iff P is true *at some world*, and $\Box P$ is true iff P is true *at every world*. Strictly speaking, $\Diamond P$ and $\Box P$ also have their truth values at worlds. If we are being precise, then, we should say that $\Diamond P$ is true at some world w_i iff P is true at some world w_j such that $w_i R w_j$, and $\Box P$ is true at some world w_i iff $\Box P$ is true at *every* world w_j such that $w_i R w_j$. Here, R is what is called an 'accessibility relation', and '$w_i R w_j$' says that w_j is *accessible from* w_i, meaning the propositions that are true at w_j bear on the relevant modal truths at w_i as per the basic definitions of '$\Diamond$' and '$\Box$'.

Once we start thinking of modal propositions as having their truth values at worlds, it becomes easier to make sense of nested modal propositions: just as an ordinary modal proposition is true at a world iff the proposition within the scope of the modal operator is true at some (every) world in the domain, so a nested modal proposition is true at a world iff the modal proposition within the scope of the *outer* modal operator is true at some (every) world in the domain. In each case, the domain just depends on the type of modality expressed by the modal operator in wide scope. For example, the domain of relevance to a nested modal proposition where the outer modal operator expresses the metaphysical modality is just the domain of metaphysically possible worlds - i.e. the set of worlds whose true propositions are metaphysically possible. After all, this is the domain over which the modal operator ranges when interpreted as a quantifier.

But here we seem to have a problem. If modal propositions like 'It is metaphysically possible that *P*' have their truth values at worlds, then from the perspective of which world do we assess the metaphysical possibility of propositions when delimiting the domain of metaphysically possible worlds? The answer is (of course) the actual world - one *could* model a counterfactual notion of metaphysical possibility from the perspective of some other world, but a counterfactual notion is a false notion, and that is not what we are interested in here. The actual world is therefore our primary *world of evaluation* for the metaphysical modality. However, this does not mean there is something special about it - it is just that, as inhabitants of the actual world, the actual world is the one whose truths we are most interested in.

Each modality comes not only with its own domain of worlds, but also with its own accessibility relation between worlds. As Lewis puts it:

> Necessity of a certain sort is truth at all possible worlds that satisfy a certain restriction. We call these worlds *accessible*, meaning thereby simply that they satisfy the restriction associated with the sort of necessity under consideration. Necessity is truth at all accessible worlds, and different sorts of necessity correspond to different accessibility restrictions. (D. K. Lewis, 1973, pp. 4-5)

Moreover, different accessibility relations can be characterised in different ways. For example, if R is reflexive, then (within the relevant domain) every world is accessible from itself: $\mathrm{w}_i\mathrm{R}\mathrm{w}_i$. If R is symmetric, then (within the relevant domain) every world is accessible from another world iff the other world is accessible from it: $\mathrm{w}_i\mathrm{R}\mathrm{w}_j$ iff w_jRw_i. And if R is transitive, then (within the relevant domain) every world that is accessible from another world is also accessible from the worlds which have access to that other world: if $\mathrm{w}_i\mathrm{R}\mathrm{w}_j$ and $\mathrm{w}_j\mathrm{R}\mathrm{w}_k$, then $\mathrm{w}_i\mathrm{R}\mathrm{w}_k$. If an accessibility relation has all three of these properties, then (within the relevant domain) every world is accessible from every world, including itself. In such cases, we say that the accessibility relation is an *equivalence* relation.

To fully grasp what is meant by this, it will be helpful to consider an example of an accessibility relation that is not an equivalence relation. Let '$\Diamond$' and '$\Box$' be interpreted in terms of a deontic modality, so that '$\Diamond P$' means that *P* is morally permissible, and '$\Box P$' means that *P* is morally obligatory. Both statements can be modelled using the possible worlds calculus in the usual way: $\Diamond P$ is true at w_i iff *P* is true at some world w_j such that $\mathrm{w}_i\mathrm{R}\mathrm{w}_j$, and $\Box P$ is true at w_i iff *P* is true at every world w_j such that

$w_i R w_j$. Let us assume that morality is *universal*, meaning that what is permissible at one metaphysically possible world is permissible at every metaphysically possible world. This in turn means that exactly the same set of worlds is morally accessible from every metaphysically possible world: namely, the set of worlds at which only morally permissible things happen. Let us call these the 'morally perfect' worlds. Some worlds which have access to the morally perfect worlds are not themselves morally perfect (including, sadly, the actual world); hence, this deontic accessibility relation is not reflexive. Nor is it symmetric, since morally perfect worlds do not have access to morally imperfect worlds. It *is* transitive: any world that is accessible from a morally perfect world is also accessible from those worlds that have access to morally perfect worlds. But the fact that it is neither reflexive nor symmetric means that it is not an equivalence relation.

What about the metaphysical accessibility relation? Lewis himself does not invoke an accessibility relation for the metaphysical modality.[1] As per the quote above, accessibility relations are a way of capturing *restrictions* in quantification, and Lewis sees metaphysical modal operators as *un*restricted quantifiers; that is, he sees them as quantifying over everything that exists. Given this, there is no need for a metaphysical accessibility relation on Lewis's view.

Why does Lewis think the metaphysical modality is unrestricted? The answer is that, even on Lewis's view, nothing metaphysically impossible exists. In other words, while some worlds have different laws of nature, no worlds have different laws of metaphysics. If one is a modal realist like Lewis, then the concept of metaphysical possibility just collapses into the concept of existence. He writes:

> [O]ther worlds are of a kind with this world... The difference between this and the other worlds is not a categorical difference.
>
> Nor does this world differ from the others in its manner of existing. I do not have the slightest idea what a difference in manner of existing is supposed to be. (D. K. Lewis, 1986a, p. 2)

However, I think it is preferable not to follow Lewis in avoiding the topic of metaphysical accessibility. To do so is to presume that there is no distinction between the metaphysical modality and the logical modality - while it may not be obvious how to distinguish these modalities, we should not presume that it cannot be done. Addition-

[1] He does invoke a *counterpart* relation: a relation that holds, not between worlds, but between *objects* in different worlds (see his (1968, 1986a)). More specifically, it holds between objects of sufficient similarity to be loosely described as 'the same', despite belonging to different worlds and (hence) being non-identical. Unlike the metaphysical accessibility relation, Lewis's counterpart relation is 'not, in general... an equivalence relation' (1968, p. 115), because it is not transitive: an object that is sufficiently similar to *x*'s counterpart may not be sufficiently similar to *x* itself. But one should be careful not to mistake Lewis's rejection of transitivity in that context as a rejection of transitivity in *this* context.

ally, there is some doubt as to whether unrestricted quantification is even coherent (see e.g. Fine (2006)). For these reasons, let us posit a metaphysical accessibility relation. The question is: What kind of accessibility relation does the metaphysical modality have?

The simplest view of any modality is one according to which the relevant modal propositions are true at one world in the domain iff they are true at every world in the domain. In other words, the simplest view of any modality is one according to which all worlds within the domain are equivalent with regard to the truth values of the relevant modal propositions. For the metaphysical modality, this means that what is metaphysically possible at the actual world is metaphysically possible at every metaphysically possible world. Now, unlike the set of morally perfect worlds, the set of metaphysically possible worlds is centred on the actual world (it includes the actual world even if it includes no others), and given this, we can assume that the metaphysical accessibility relation is just the *universal* relation: a special kind of equivalence relation according to which all members of a set are related to all members, including themselves. In this context, what that means is that all metaphysically possible worlds are accessible to all metaphysically possible worlds.

In general, our starting assumption for *any* type of possibility (as opposed to any type of *permissibility*) ought to be that the relevant accessibility relation is the universal relation. Of course, there is nothing to stop one from *inventing* a type of possibility that comes with a different kind of accessibility relation, but in this context, we are interested only in modalities that are expressed in natural language, and we should complicate our models of these modalities only if we find justification for doing so upon consideration of the consequences of not doing so. For these reasons, let us take it as given that the metaphysical accessibility relation is the universal relation.

It is worth noting that, in this context, accessibility via the universal relation is effectively just unrestricted quantification by another name, since no metaphysically impossible worlds exist. Given the conception of other possible worlds as 'of a kind with this world' (D. K. Lewis, 1986a, p. 2), this is exactly what we ought to expect for metaphysical accessibility. On Lewis's view, there is nothing special about the actual world compared to other worlds - it is our primary world of evaluation only because it is the world at which we happen to exist.

We are now in a good position to use the possible worlds calculus to interpret metaphysical nested modal propositions in a way that explains why their assertions seem infelicitous. Consider again the statement 'It is metaphysically possible that P is metaphysically possible'. In symbols, we can write this as '$\Diamond\Diamond P$', where both operators express the metaphysical modality. Applying the usual definition of '$\Diamond$' allows us to say that $\Diamond\Diamond P$ is true at some world w_i iff $\Diamond P$ is true at some world w_j such that $\mathrm{w}_i\mathrm{R}\mathrm{w}_j$, where R is the metaphysical accessibility relation. But under the supposition that R is the universal relation, every metaphysically possible world is accessible from every metaphysically possible world. This means that $\Diamond\Diamond P$ is true at *any* world in the domain iff $\Diamond P$ is true at *any* world in the domain. Hence, $\Diamond\Diamond P$ and $\Diamond P$ have exactly the same truth conditions. This explains why the assertion of 'It is metaphysically possible that P is metaphysically possible' seems infelicitous: one could just say 'P is metaphysically possible' instead; hence, by asserting the longer proposition, one violates the Gricean maxim of Manner, 'Be brief'. (1967b, p. 27) (See section 4.1 for more details.)

Now consider the statement 'It is metaphysically necessary that P is metaphysically necessary.' In symbols, this can be written '$\Box\Box P$'. Applying the usual definition of '$\Box$' allows us to say that $\Box\Box P$ is true at some world w_i iff $\Box P$ is true at every world w_j such that $w_i R w_j$. But under the supposition that R is the universal relation, any metaphysical modal proposition is true at *some* world in the domain iff it is true at *every* world in the domain. Hence, $\Box\Box P$ is true at every metaphysically possible world iff $\Box P$ is true at every metaphysically possible world. So $\Box\Box P$ turns out to have exactly the same truth conditions as $\Box P$. This explains why the assertion of 'It is metaphysically necessary that P is metaphysically necessary' seems infelicitous - one could just say 'P is metaphysically necessary' instead.[2]

We have now covered our two initial examples: $\Diamond\Diamond P$ and $\Box\Box P$. But what about $\Box\Diamond P$ and $\Diamond\Box P$? For similar reasons, if the relevant accessibility relation is the universal relation, then these turn out to be equivalent to $\Diamond P$ and $\Box P$ respectively. In other words, if the relevant accessibility relation is the universal relation, then we can interpret simple nested modal propositions *by ignoring the outer modal operator*. This rule works even when there is a negation sign to take into consideration. For example, $\Box\neg\Diamond P$ has the same truth conditions as $\neg\Diamond P$, and $\neg\Diamond\Box P$ has the same truth conditions as $\neg\Box P$.

Equipped with this 'ignore the outer operator' rule, it becomes slightly easier to understand the consequences of a metaphysically strict analysis for conditionals where the consequent features a metaphysical modal operator. Let us say that a conditional with a metaphysically necessary consequent is 'trivially' true. Now consider the following conditional:

> $\mathbf{G} \to \Box\mathbf{S}$ *If I was born in Glasgow, then it is metaphysically necessary that I was born in Scotland.*

This conditional is trivially true iff $\Box\Box\mathbf{S}$ is true. And given the supposition that the metaphysical accessibility relation is the universal relation, this latter proposition is true iff $\Box\mathbf{S}$ is true. Hence, $\mathbf{G} \to \Box\mathbf{S}$ has the same conditions for trivial truth as $\mathbf{G} \to \mathbf{S}$.

[2] For what it is worth, this also validates a principle of modal logic known as *Becker's Principle*: $\Box P \supset \Box\Box P$.

Of course, we are interested in more than just *trivial* truth conditions. Moreover, on Williamson's view, the 'would' operator expresses circumstantial necessity, not metaphysical necessity. In order to explain the consequences of combining a Williamsonian view of 'would' with a metaphysically strict analysis of 'would'-conditionals, we need to know how to understand *complex* nested modal propositions. That will be the subject of the next section.

6.3 Complex Nested Modality

Whilst assertions of simple nested modal propositions often seem to be infelicitous, assertions of complex nested modal propositions often seem to be felicitous. Suppose, for example, that you are a physicist who entertains post-Einsteinian theories. You may felicitously respond 'Possibly' when asked 'Is it physically possible to travel faster than the speed of light?'. The most natural interpretation of such a response is as a complex nested modal proposition where the outer operator expresses epistemic possibility.

To keep track of the different modalities expressed by modal operators, let us attach subscripts according to the initial letters of the names of those modalities. For example, '$\Diamond_e \Diamond_p P$' can be understood as meaning 'It is *epistemically* possible that P is *physically* possible'. One might wonder whether we can use the 'ignore the outer operator' rule to simplify these statements, just as we did with the simple nested modal propositions. However, this rule does not apply to complex nested modal propositions. To see this, consider the formula '$\Box_m \Box_c P$', where this means 'It is metaphysically necessary that it is circumstantially necessary that P.' Assuming that the metaphysical accessibility relation is the universal relation, a metaphysical modal proposition like this one is true at *any* metaphysically possible world iff it is true at *every* metaphysically possible world. With this in mind, applying the definition of '$\Box$' gives us:

> $\Box_m \Box_c P$ is true at *every* metaphysically possible world iff $\Box_c P$ is true at every metaphysically possible world.

Now, if the circumstantial modality were identical to the metaphysical modality, then we could ignore the outer modal operator completely, and say:

> $\Box_m \Box_c P$ is true at every metaphysically possible world iff P is true at every *circumstantially* possible world.

However, the circumstantial and metaphysical modalities are not identical. In most contexts, the set of circumstantially possible worlds is a proper subset of the set of metaphysically possible worlds, and in any such context, all we can say is:

> $\Box_m \Box_c P$ is true at every metaphysically possible world *only if* P is true at every circumstantially possible world.

In other words, even if P is true at every circumstantially possible world, it might be that $\neg P$ is circumstantially possible from the perspective of some circumstantially *im*possible world. Thus, we cannot ignore the outer modal operator completely.

Nonetheless, let us see what happens when we combine a Williamsonian view of 'would' with a metaphysically strict analysis of 'would'-conditionals. Supposing Williamson is right that 'would' is a circumstantial necessity operator that can feature in unconditional propositions, let us represent unconditional 'would'-statements as '$\Box_c D$', where D is the "difference" between the 'would'-statement and the 'would'.[3] For example, in (4), D is the proposition expressed by 'Loretta does not betray a friend.' With this in mind, consider the following example of Williamson's (2020, p. 171):

> (5) *If Loretta were loyal, then she would not betray a friend.*

Let us represent (5) as '$A > \Box_c D$', where '$>$' is taken to express metaphysically strict implication. Since 'Loretta were loyal' is ungrammatical, let us assume that A is just the proposition expressed by 'Loretta is loyal'. As Williamson notes, 'were' is anaphoric on 'would' (2020, p. 171): it signals to us that 'would' is coming in the consequent. Now, if '$>$' is metaphysically strict, then $A > \Box_c D$ can be analysed further as $\Box_m(A \supset \Box_c D)$. Hence, on this interpretation of 'would'-conditionals, (5) effectively says 'It is metaphysically necessary that: either Loretta is not loyal or, in all contextually relevant worlds, she does not betray a friend.'

It is not totally implausible that this is what (5) says. But the interpretation of 'would'-conditionals that has emerged attributes to them quite an unwieldy logical structure, one which many competent users of 'would'-conditionals may struggle to intuitively grasp. Indeed, even *trivial* truth is quite difficult to grasp: (5) is trivially true iff it is metaphysically necessary that 'Loretta does not betray a friend' is circumstantially necessary. If we could simply ignore the outer operator, then this would be easy to understand, but we cannot. I think a more elegant and intuitive view of 'would'-conditionals is available. In the next section, I explain that view.

6.4 The 'Would' Operator Revisited

Williamson combines his view of 'would' with a view according to which conditionals are material. One might think, therefore, that he takes (5) to be synonymous with:

> (5a) Either Loretta is loyal or, in all contextually relevant worlds, she does not betray a friend.

On the contrary, he writes:

[3] I would use the word 'complement', but the letter 'C' is already taken.

> [T]he natural analysis of [(5)] according to the proposed schema assigns 'would' wide scope with respect to 'if':
>
> [(5b)] Would(if Loretta is loyal, Loretta does not betray a friend).
>
> (Williamson, 2020, p. 171)

I think Williamson is right that the 'would' operator in (5) takes both the antecedent and consequent within its scope. However, unlike Williamson, I think this is true of 'would'-conditionals *in general*, as I explain.

Compare the following: $A \supset \Box C$ and $\Box(A \supset C)$. These formulae say different things: the first says that either $\neg A$ is true or C is necessary, whereas the second says that the material conditional as a whole is necessary. To confuse these formulae is to commit a common modal scope fallacy. However, it is not hard to see why it is common, since many natural language conditionals seem to say exactly the same thing no matter where we place the modal operator. For example, compare the following:

(6a) *If Oswald did not kill Kennedy, then necessarily: someone else did.*

(6b) *Necessarily: if Oswald did not kill Kennedy, then someone else did.*

In any non-philosophical context, these statements are interchangeable. This is perhaps aided by the fact that the scope of the modal operator is ambiguous in the following:

(6c) *Necessarily: someone else killed Kennedy if Oswald did not.*

Even though the 'would' operator in a 'would'-conditional appears to take only the consequent within its scope, it may be that there is a better way to represent the 'would'-conditional's logical structure. Since it is a common fallacy to confuse $A \supset \Box C$ with $\Box(A \supset C)$, it may be that what people typically mean when they use a 'would'-conditional is a proposition where the 'would' operator takes both the antecedent *and* consequent within its scope. Indeed, this helps to explain the intuitive pull towards treating D as the 'would'-conditional's consequent: since the 'would' operator takes both the antecedent and consequent within its scope, it is not a part of either, so D is the whole consequent after all.

In addition to this, I think we have independent reason to believe that the 'would' operator is a metaphysical necessity operator, not a circumstantial one. Williamson disagrees, writing that 'in an ordinary context (4) does not claim that it is metaphysically impossible for Loretta to betray a friend: she could have been brought up in very different circumstances and acquired a different character.' (2020, p. 170) But note that even Williamson's view is consistent with the possibility that in *some* contexts, (4) *does* claim that it is metaphysically impossible for Loretta to betray a friend. The reason Williamson thinks there are some contexts in which (4) does not claim this is just that, for some particularly loyal referents of the name 'Loretta', (4) seems to be true. Yet we should not be wedded to the idea that (4) is true of such people. The price we pay for being wedded to that idea is context-sensitivity as per the definition of 'circumstantial modality', and that is a hefty price indeed. If unconditional 'would'-statements are context-sensitive, then many of them are true some of the time, but we have no principled way of determining *when* they are true, because we have no principled way of determining contextual relevance.

One reason to doubt my contention that 'would' is a metaphysical necessity operator is that its accessibility relation does not seem to be an equivalence relation. On this front, Williamson writes:

> [W]hen we are developing a scenario in which a tiger enters the room, we may truly utter [(7)], even though in the actual world Francis is not frightened:
>
> [(7)] Francis would be frightened.
>
> ...R need not be a reflexive relation; as noted with example [(7)], 'would(A)' does not always imply A. (Williamson, 2020, pp. 170-174)

However, this sort of counterexample only works on the assumption that (7) is true in this situation, and I do not think there is any compelling reason to believe that (7) is true in this situation, as I will explain.

It would be quite a leap to conclude on the basis of statements like (4) that the traditional thesis (i.e. the thesis that unconditional 'would'-statements are the consequents of implicit conditionals) is false for *all* 'would'-statements. Even if one can make good sense of (4) as an unconditional 'would'-statement, it may still make more sense to interpret other 'would'-statements as the consequents of implicit conditionals. In particular, I think it is clear that (7) should be interpreted as the consequent of:

(8) *If a tiger were to enter the room, Francis would be frightened.*

Interpreted as such, we can evaluate the assertion of (7) against KNAC (see section 4.4) on the supposition that this norm governs the assertion of natural language conditionals in general. If this supposition is correct, and if (7) is the consequent of an implicit conditional, then the assertion of (7) is proper only if one knows that the corresponding material conditional is true at all worlds of interest to the participants in the relevant conversation. Since it is possible for an asserter to know that 'Either a tiger does not enter the room or Francis is frightened' is true at all worlds of interest even if it is not true at all metaphysically possible worlds, (7) may be asserted even if the implicit conditional of which it is the consequent is false, and indeed even if (7) itself is false.

Since (7) does not need to be true to be assertable, I do not think we have a compelling reason to believe that it is true in the situation described by Williamson above.

With this said, let us now see what happens if we combine a metaphysically strict analysis of 'would'-conditionals with a view of 'would' as a metaphysical necessity operator that takes both the antecedent and consequent of a 'would'-conditional within its scope. On the resultant interpretation, a 'would'-conditional effectively says 'It is metaphysically necessary that it is metaphysically necessary that either $\neg A$ or C'. Since this is a simple nested modal proposition where both operators express the metaphysical modality, it may be simplified using the 'ignore the outer operator' rule to give us: 'It is metaphysically necessary that either $\neg A$ or C'. For example, (5) can be understood as saying 'It is metaphysically necessary that either Loretta is not loyal or she does not betray a friend.'

This interpretation of 'would'-conditionals is of course one according to which they have exactly the same truth conditions as their corresponding indicative conditionals. In effect, the 'would' operator makes no difference to the truth conditions of the conditional. If this is right, then the difference between 'would'-conditionals and indicative conditionals is not semantic but pragmatic. As argued in section 5.7, a difference of assertability can arise from the fact that 'would'-conditionals sometimes convey a belief that everyone in the conversation knows that the antecedent of the conditional is false. Nonetheless, the proponent of the metaphysically strict analysis of indicative conditionals can reasonably maintain that the truth conditions of a 'would'-conditional are exactly the same as its corresponding indicative conditional.

One might object that this interpretation of 'would'-conditionals makes too many 'would'-conditionals false. But just as KNAC can protect the metaphysically strict analysis of indicative conditionals against this objection (see chapter 4), so it can protect a metaphysically strict analysis of 'would'-conditionals against this objection. Moreover, it is much less controversial to believe in the widespread falsehood of 'would'-conditionals than it is to believe in the widespread falsehood of indicative conditionals. Consider e.g. Hájek's (2014) paper 'Most Counterfactuals Are False', in which he presents the following argument:

> [S]uppose for now that coin-tossing is a chancy business: as a coin is tossed, it is a genuinely indeterministic matter whether it lands heads or tails... Here is a coin that in fact will never be tossed. Consider the counterfactual: 'If the coin were tossed, it would land heads'... There is no particular way that this chancy process *would* turn out, were it to be initiated... To think that there is a fact of the matter of how the coin would land is to misunderstand chance. (Hájek, 2014, pp. 6-7)

Hájek uses the coin-flip example to goad our intuitions, but he thinks the same argument applies to the vast majority of 'would'-conditionals used in everyday contexts. And while Hájek's conclusion strikes some philosophers as absurd, it strikes others as perfectly reasonable. He writes:

> I have presented my arguments to many philosophical audiences over a number of years... One common reaction, only slightly caricatured, is that *I have lost my philosophical marbles*... [But a]nother common reaction is that *these arguments are entirely sound and persuasive*—end of story. (Hájek, 2014, pp. 1-2)

So, when responding to the objection that a metaphysically strict analysis of 'would'-conditionals makes too many conditionals false, I can appeal not only to KNAC, but also to the fact that I have a companion in guilt.

Another possible objection concerns some purported principles of conditional logic:

Conditional Excluded Middle (CEM) $(P > Q) \lor (P > \neg Q)$,

Conditional Non-Contradiction (CNC) $\neg((P > Q) \land (P > \neg Q))$.

One might object to a metaphysically strict analysis of 'would'-conditionals on the basis that it validates neither CEM nor CNC, despite their intuitive appeal.[4] The reasons for this are just the same as the reasons that the metaphysically strict analysis of indicative conditionals does not validate either of these principles. CEM says that a conditional and its contrary cannot both be false, but if the metaphysically strict analysis of 'would'-conditionals is true, then a 'would'-conditional may be false just because its corresponding material conditional is metaphysically contingent, in which case its contrary will be false as well. CNC says a conditional and its contrary cannot both be *true*, but if the metaphysically strict analysis of 'would'-conditionals is true, then a 'would'-conditional may be *vacuously* true if its antecedent is metaphysically impossible, in which case its contrary will be vacuously true as well.

The invalidation of these principles may seem unappealing, but as noted in chapter 2, Lewis's analysis of counterfactuals also invalidates CEM, and the material interpretation of indicative conditionals also invalidates CNC. In each case, then, I have a companion in guilt.

6.5 The 'Might' Operator

As well as 'would'-conditionals, there are also 'might'-conditionals. For example:

[4] There are two caveats to this claim that are worth mentioning. The first is that the metaphysically strict analysis validates CEM if strict necessitarianism is true (see section 3.4). The second is that it (also) validates CNC if CNC is restricted to conditionals with metaphysically possible antecedents.

(9) *If Loretta were loyal, then she* might *betray a friend.*

In order for a metaphysically strict analysis of 'would'-conditionals to be credible, it ought to fit with a plausible analysis of 'might'-conditionals. Thankfully, the interpretation of 'would'-conditionals defended in this chapter points to a highly plausible analysis of 'might'-conditionals.

'Might'-conditionals are often represented using the '$\Diamond\!\!\rightarrow$' symbol, which is related to '$\Box\!\!\rightarrow$' as follows:

$$P \Diamond\!\!\rightarrow Q \equiv \neg(P \Box\!\!\rightarrow \neg Q),$$
$$P \Box\!\!\rightarrow Q \equiv \neg(P \Diamond\!\!\rightarrow \neg Q).$$

The upshot of this is that a 'might'-conditional is the contradictory of its contrary 'would'-conditional, and vice versa. This coheres with the intuition that (9) contradicts (5). Indeed, contradicting contrary 'would'-conditionals seems to be the primary purpose of 'might'-conditionals. (Unlike 'would'-conditionals, 'might'-conditionals are not very useful.) According to Bennett, the above equivalences are a condition 'on any reasonable understanding of subjunctive conditionals'. (2003, p. 192) Indeed, Lewis (1973, p. 2) simply *defines* 'might'-conditionals this way.[5] In what follows, I take these equivalences as given. The view of 'might'-conditionals that emerges is one according to which they are not really conditionals at all: they are statements of metaphysical *possibility* as opposed to statements of metaphysical *necessity*.

Given the dual roles of 'would'-conditionals and 'might'-conditionals, it is desirable to treat 'would' and 'might' analogously (*mutatis mutandis*). For example, if we treat 'would' as a metaphysical modal operator that can feature in unconditional propositions, we ought also to treat 'might' as a metaphysical modal operator that can feature in unconditional propositions. But whereas 'would' is a *necessity* operator, 'might' is a *possibility* operator. For example, if I were to say (10), I would convey that Loretta betrays a friend in *some* metaphysically possible world(s):

(10) *Loretta might betray a friend.*

Dealing with unconditional 'might'-statements is fairly straightforward. Dealing with 'might'-conditionals, on the other hand, is more complicated. If we treat 'would' as taking both the antecedent and consequent of a 'would'-conditional within its scope, we should also treat 'might' as taking both the antecedent and consequent of a 'might'-conditional within its scope. But whereas a 'would'-conditional is ultimately a material conditional within the scope of the 'would' operator, a 'might'-conditional is surely *not* a material conditional within the scope of the 'might' operator. On such a view, (9) would effectively say 'It is metaphysically possible that either Loretta is not loyal or she

[5] See Bigelow and Pargetter (1990, p. 103) for another endorsement of these equivalences.

betrays a friend'. The mere metaphysical possibility of Loretta's not being loyal would suffice to make (9) true. That is absurd. The truth conditions of 'might'-conditionals are easily satisfied, but they are not *that* easily satisfied.

Thankfully, there is independent reason to think that 'might'-conditionals have a different logical structure. As above, the primary purpose of a 'might'-conditional is to contradict its contrary 'would'-conditional. However, $\Diamond(P \supset Q)$ does not contradict $\Box(P \supset \neg Q)$; rather, $\Diamond(P \wedge Q)$ does.[6]

If a 'might'-conditional is a conjunction within the scope of a metaphysical possibility operator, then (9) effectively says 'It is metaphysically possible that Loretta is loyal *and* she betrays a friend'. I think it is highly plausible that this is indeed what (9) says. On this view, the truth conditions of 'might'-conditionals are easily satisfied, but not to the point of absurdity. Some might object that the 'might' operator should be understood as ranging over only those metaphysically possible worlds that are contextually relevant. That is indeed the sensible view if one also takes the 'would' operator to range over only those metaphysically possible worlds that are contextually relevant. But I hope to have shown that one can defend a more elegant view according to which both operators express the metaphysical modality. Such a view is more elegant not only because it avoids context-sensitivity, but also because it coheres with the metaphysically strict analysis of indicative conditionals to form a unified analysis of natural language conditionals in general.

It is worth pointing out that if a metaphysically strict analysis of natural language conditionals is right, then 'might'-conditionals may also be used to contradict contrary indicative conditionals. Suppose, for example, that I am a nutty conspiracy theorist who is entertaining the idea that Kennedy's assassination was faked: not only was Oswald framed, but Kennedy lived out the rest of his days in Russia before dying of natural causes. If you were to say 'If Oswald did not kill Kennedy, then someone else did', I might naturally contradict you by saying 'If Oswald did not kill Kennedy... then it *might* be that no one else did either!'

Finally, one might worry that what I have said about 'might'-conditionals in this section contradicts what I have said about the number of logical types of natural language conditional in chapter 5. In that chapter, I argued that there is just *one* logical type of natural language conditional, yet here I am saying that 'might'-conditionals have a completely different logical structure to both indicative conditionals and 'would'-conditionals. I do not see these statements as contradictory, because I think the view of 'might'-conditionals defended in this section is one according to which they are not really conditionals at all; rather, they are the *negations* of conditionals. After all, $\Diamond(P \wedge Q)$

[6] To see this, note that $\Diamond(P \wedge Q)$ is equivalent to $\Diamond\neg(P \supset \neg Q)$, since $P \supset \neg Q$ is false iff P and Q are both true. Note also that $\Diamond\neg(P \supset \neg Q)$ is equivalent to $\neg\Box(P \supset \neg Q)$, given the interdefinability of '$\Box$' and '$\Diamond$'. Hence, $\Diamond(P \wedge Q)$ is equivalent to $\neg\Box(P \supset \neg Q)$.

is equivalent to $\neg\Box(P \supset \neg Q)$. However, if one insists on thinking of these propositions as conditionals, then I am willing to concede that there are two different logical types of natural language conditional: non-negated natural language conditionals and negated natural language conditionals.

6.6 Conclusion

In this chapter, I have argued that the metaphysically strict analysis of indicative conditionals may be extended to cover 'would'-conditionals as well. I do not expect to have persuaded the reader that this is the correct analysis of 'would'-conditionals; rather, my aim has been to strengthen the metaphysically strict analysis of indicative conditionals by showing that it coheres with a defensible analysis of 'would'-conditionals according to which they are also metaphysically strict. Finally, I have argued that this analysis of 'would'-conditionals itself coheres with a defensible analysis of 'might'-conditionals according to which they are not really conditionals at all. In the next chapter, I show that this unified metaphysically strict analysis allows us to draw a connection between conditionals and conditional probability.

Chapter 7

Conditionals and Conditional Probability

Intuitively, there is *some* sort of connection between conditionals and conditional probability. As Ramsey once wrote:

> If two people are arguing 'If p will q?' and are both in doubt as to p, they are adding p hypothetically to their stock of knowledge and arguing on that basis about q... We can say they are fixing their degrees of belief in q given p. (Ramsey, 1931a, p. 247, footnote 1)

Degrees of belief can be thought of as subjective probabilities (i.e. probabilities from the viewpoint of a particular person), and the probability of one thing *given* another is a conditional probability; hence, Ramsey is suggesting a connection between conditionals and conditional probability.

The idea that there is some such connection between conditionals and conditional probability is so intuitive as to be almost *platitudinous*. Indeed, Ramsey's footnote is one of the most oft-quoted texts in the literature.[1] The deliberative procedure described by Ramsey (wherein one assesses a conditional by fixing one's degree of belief in its consequent given its antecedent) has come to be called 'the Ramsey test', and according to Bennett, '[t]he literature on indicative conditionals is a parade of attempts to explain why the Ramsey test is a valid criterion for their acceptability.' (2003, p. 30) One of the most prominent floats in that parade is Adams' Thesis, according to which the 'justified assertability' of an indicative conditional corresponds to the conditional probability of its consequent given its antecedent. (Adams, 1965, *passim*) This is one way to flesh out the intuitive connection between conditionals and conditional probability, and it has proven to be highly influential. However, even if Adams' Thesis is right, we

[1] It can be found, for example, in Edgington (1995, p. 264), Read and Edgington (1995, p. 47), Bennett (2003, p. 28), and Chalmers and Hájek (2007, p. 170), to name just a few.

might still think there is a more fundamental connection that does not involve 'justified assertability'. The question is: What might that more fundamental connection be?

Of the candidate connections, the most obvious is just that the probability of a conditional is equal to the conditional probability of its consequent given its antecedent. This has come to be known as 'Stalnaker's Thesis' after Stalnaker (1970b) built a system of logic around the idea, but it has also been proposed by both Jeffrey (1964) and Ellis (1969). Unfortunately, as we will see in section 7.8, Stalnaker's Thesis faces compelling counterarguments based on Lewis's (1976) triviality results - indeed, they are so compelling that even Stalnaker has given up on Stalnaker's Thesis. In sum, it remains a matter of debate how to account for the intuition that there is some sort of connection between conditionals and conditional probability.

In this chapter, I explain how we can flesh out the aforementioned intuitive connection if the unified strict analysis of conditionals that I defend is true. According to that strict analysis, conditionals are true iff the conjunction of their antecedent and their consequent's negation is a metaphysical impossibility. Given this, and given the right kind of probability function, we can draw a connection between conditionals and conditional probability which we might as well call 'McCardel's Thesis' (for want of a less pretentious name). This connection is as follows:

McCardel's Thesis $A > C$ is true iff $\mathrm{P}(C|A) = 1$.

This helps us to make sense of the Ramsey test as a way of evaluating conditionals: if one's degree of belief in C given A is 1, then one ought to accept the conditional $A > C$. As we will see in chapter 8, it also helps us to make sense of other philosophical issues. Moreover, this connection is not viable on anything other than a strict analysis of conditionals, and it is most plausible given a metaphysically strict analysis in particular. This constitutes a big part of the argument in favour of the metaphysically strict analysis.

The defence of McCardel's Thesis hinges on what is meant by 'the right kind of probability function'. For that reason, most of this chapter is devoted to carving out different types of probability, and explaining key concepts and rules in the theory of probability. In particular, I argue that a Lewisian conception of objective chance fits the bill. According to this conception, chance is rational subjective probability conditional on history plus history-to-chance conditionals. In sections 7.1 and 7.2, I explain what is meant by rational subjective probability. I then explain the concepts of conditionalisation and objective probability, as well as Lewis's Principal Principle, all of which must be understood in order to make sense of his conception of objective chance. This allows us to make use of a constraint known as *regularity*, which I discuss in section 7.5. With this constraint in play, I prove that McCardel's Thesis is true *if* the unified analysis of conditionals that I defend is also true. In section 7.8, I ward off counterarguments of the sort that plague Stalnaker's Thesis, and in section 7.9, I ward off counterarguments posed by Edgington against a similar view. Finally, I conclude.

7.1 The Rules of Probability

In his *Logical Foundations of Probability* (1950), Carnap points out that we sometimes mean quite different things by the word 'probability', and that authors on the subject have in some cases attempted to explicate quite different notions:

> When we look at the formulations which the authors themselves offer in order to make clear which meanings of 'probability' they intend to take as their explicanda, we find phrases as different as 'degree of belief', 'credibility', 'degree of reasonable expectation', 'degree of possibility', 'degree of proximity to certainty', 'degree of partial truth', 'relative frequency', and many others. This multiplicity of phrases shows that any assumption of a unique explicandum common to all authors is untenable. (Carnap, 1950, p. 24)

He goes on to argue that these different phrases reveal two fundamentally different explicanda: degree of confirmation (which he calls '$\textit{probability}_1$'), and relative frequency (which he calls '$\textit{probability}_2$').

Whether or not Carnap is right about this, it is important to remember that each theory of probability is an attempt at explicating a pre-theoretical notion, and there is no guarantee that we all have the same pre-theoretical notion in mind. By far the most well-known and influential attempt at explicating probability comes from Kolmogorov. Indeed, his explication has been so successful that the axioms he suggested are nowadays considered to be *the* axioms of probability theory. He writes:

> Let E be a collection of elements... which we shall call *elementary events*, and $\mathcal{F}$ be a set of subsets of E; the elements of the set $\mathcal{F}$ will be called *random events*.
>
> I. *$\mathcal{F}$ is a field of sets.*
>
> II. *$\mathcal{F}$ contains the set E.*
>
> III. *To each set A in $\mathcal{F}$ is assigned a non-negative real number* $\mathrm{P}(A)$. *This number* $\mathrm{P}(A)$ *is called the probability of the event A.*
>
> IV. $\mathrm{P}(E)$ *equals* 1.
>
> V. *If A and B have no element in common, then*
>
> $$\mathrm{P}(A + B) = \mathrm{P}(A) + \mathrm{P}(B)$$

> A system of sets, $\mathcal{F}$, together with a definite assignment of numbers P(A), satisfying axioms I-V, is called a *field of probability*. (Kolmogorov, 1956, p. 2)

Even though Kolmogorov lists five axioms in the passage just quoted, it is the last three axioms in particular that have been taken up as definitive of probability. Let us call these 'the non-negativity axiom,' 'the normalization axiom,' and 'the additivity axiom' respectively. The non-negativity axiom is easy to understand: P gives us real numbers in the unit interval (between and including 0 and 1). The normalization axiom is also fairly straightforward: it means that the probability that *some* event in the total set will occur is 1. The additivity axiom is a little more complicated: it means that the probability of the union of any disjoint subsets of the total set is equal to the sum of the probabilities of each of those subsets. More intuitively: the probability that at least one of two mutually exclusive events will occur is just the probability of the first event plus the probability of the second.

The Kolmogorov axioms are an attempt to capture our intuitions about probability in as efficient a way as possible, the idea being that any intuitions not explicitly captured by the axioms will nonetheless be captured by theorems that can be derived from them. For instance, it follows from these axioms that the probability of the empty set P($\varnothing$) is equal to 0, and that if B is a subset of A, then P(A) ≥ P(B). As above, these axioms have become so firmly entrenched in the literature that they are often taken to be *definitive* of probability. And of course, it is the nature of mathematics that *some* set of axioms will become definitive. As Kolmogorov writes:

> The theory of probability, as a mathematical discipline, can and should be developed from axioms in exactly the same way as Geometry and Algebra. This means that after we have defined the elements to be studied and their basic relations, and have stated the axioms by which these relations are to be governed, all further exposition must be based exclusively on these axioms... (Kolmogorov, 1956, p. 1)

Nonetheless, we should not lose sight of the fact that Kolmogorov was attempting to explicate a pre-theoretical notion of probability. His attempt may have been highly successful, but that does not mean there are no better explications out there. For instance, it may be that there are constraints on probability that cannot be derived from the Kolmogorov axioms, or even that these axioms constrain things *too much*. Importantly, we can consider these possibilities without changing the subject from probability to something else.

As can be seen, Kolmogorov's axioms are expressed in the terminology of set theory, and the elements (or members) of the sets are *events*. However, in specifying a probability *logic*, we will take the elements to be propositions. So, where Kolmogorov writes 'P(A)' to mean the probability of the event, A, we write 'P(A)' to mean the probability that the proposition, A, is true. Every event can be captured by a corresponding proposition: the proposition that the event occurs. Moreover, this brings our interpretation in line with contemporary philosophical literature without doing any harm to the mathematics involved. As Kolmogorov writes:

> [T]he concept of a *field of probabilities* is defined as a system of sets which satisfies certain conditions. What the elements of this set represent is of no importance in the purely mathematical development of the theory of probability. . . ' (Kolmogorov, 1956, p. 1)

Here, P is a probability *function* - it takes propositions as inputs and gives as outputs real numbers in the unit interval. Alternatively, P is sometimes conceived as a probability *distribution* - it distributes numbers, assigning them to propositions. In this context, I will treat 'function' and 'distribution' as synonymous.

Corresponding to the Kolmogorov axioms are certain rules of *probability logic*. Some of these rules use words like 'necessary', 'inconsistent', 'impossible', and 'implies'. As we have seen (in chapter 3), such words are ambiguous between different types of modality. Furthermore, they are often left ambiguous by those who write about rules of probability logic (e.g. Bennett (2003, pp. 49-50)). Their interpretation in modal logic is normally in terms of *metaphysical* possibility, since modal claims are modelled in terms of possible worlds, and the existence of other possible worlds is a metaphysical thesis. This is the interpretation we will adopt here. Interpreting the rules this way is perfectly consistent with the set-theoretic language of the Kolmogorov axioms, and it allows us to draw a connection between conditionals and conditional probability on the theory of conditionals that I defend.

So, corresponding to the Kolmogorov axioms, we have:

rule 1 $0 \leq \mathrm{P}(A) \leq 1$.

rule 2 If A is (metaphysically) necessary, then $\mathrm{P}(A) = 1$.

rule 3 If A and B are (metaphysically) inconsistent, then $\mathrm{P}(A \vee B) = \mathrm{P}(A) + \mathrm{P}(B)$.

Corresponding to the two theorems that were mentioned above, we also have:

rule 4 If A is (metaphysically) impossible, then $\mathrm{P}(A) = 0$.

rule 5 If A (metaphysically) implies B, then $\mathrm{P}(A) \leq \mathrm{P}(B)$.

Many more rules can be derived still. For example, from rule 2, it follows that

rule 6 $\mathrm{P}(A \vee \neg A) = 1$, since $A \vee \neg A$ is metaphysically necessary. From rules 3 and 6, it follows that

rule 7 $\mathrm{P}(A) + \mathrm{P}(\neg A) = 1$,

since A and $\neg A$ are metaphysically inconsistent. And from rule 4, it follows that

rule 8 $\mathrm{P}(A) \geq \mathrm{P}(A \wedge B)$,

since $A \wedge B$ metaphysically implies A. And so on. Of course, it may be that versions of these rules in terms of a less permissive type of modality are *also* true, e.g. 'If A is (*physically*) impossible, then $\mathrm{P}(A) = 0$.' But for our current purposes, these versions of the rules will do.

7.2 Subjective Probability

Nowadays, it is standard to draw a distinction between subjective and objective probability.[2] A subjective probability function is one that represents a subject's view of the world: the number assigned to a proposition represents the degree to which the subject believes the proposition. Degrees of belief are also sometimes known as 'credences'. When dealing with subjective probability functions, to make it clear what type of probability function we are dealing with, I will write 'Cr' for 'credence' (instead of 'P' for 'probability'). For instance, if I am completely sure that A is true, then $\mathrm{Cr}(A) = 1$; and if I am completely sure that A is false, then $\mathrm{Cr}(A) = 0$.

Some sources assume that all subjective probability functions represent the credences of *rational* agents in particular - see e.g. Hájek (2003, p. 276). This is a perfectly reasonable assumption to make, especially if one takes the Kolmogorov axioms to be definitive of probability, because it ensures that the topic does not drift from probability functions to functions of some other type. However, I will not be making the same assumption. The rules of logic (whether probability logic or otherwise) are normative: it is (epistemically) *good* if one's credences conform to them. But it seems to me that one's credences do not cease to be subjective probabilities if they break the rules.[3] In the same vein, Bennett writes:

[2] See e.g. Bennett (2003, §20).

[3] It is well known that they *do* break the rules, thanks in part to the research of Kahneman and Tversky. Consider, for example:

> Linda is thirty-one years old, single, outspoken, and very bright... majored in philosophy... [and] deeply concerned with issues of discrimination and social justice... Which alternative is more probable? Linda is a bank teller... [or] Linda is a bank teller and is active in the feminist movement. (Kahneman, 2011, pp. 156-8)

The answer (by rule 8) is that the former is more probable than the latter, but an embarrassingly high number of people initially assign greater subjective probability to the latter proposition - about 85% to 90% of those surveyed, apparently. (2011, p. 158)

> Considered as a thesis about subjective probability, [rule 3] might be false of an individual person... But a probability *logic* is normative: it sets constraints on how people's degrees of closeness to certainty should behave and combine, as do the laws of ordinary logic and arithmetic... This is the spirit in which we must understand all the axioms and theorems. (Bennett, 2003, p. 51)

So if, for example, $\mathrm{Cr}(A) = 0.65$ and A entails B, then (by rule 3) it *should* be the case that $\mathrm{Cr}(B) \leq 0.65$; and (by rule 4) it *should* be the case that $\mathrm{P}(A \wedge (B \vee \neg B)) = 0.65$; and (by rule 7) it *should* be the case that $\mathrm{Cr}(\neg A) = 0.35$. And so on.

It will be useful to have a term to distinguish those subjective probability functions that conform to the rules of probability logic, whether or not those rules are the ones specified above. Probability functions that conform to the rules are said to be *coherent*.[4] Coherence seems to be necessary for rationality or reasonableness on the part of the subject whose credences are represented, but whether it is sufficient is contentious.[5] Even if the rules above are correct, it is not implausible that there are other constraints on rationality or reasonableness for subjective probability functions. As Hájek puts it:

> We may claim to embrace... [this] freewheeling epistemology according to which rational belief is merely a matter of probabilistic 'coherence'. But in our hearts, we know that rationality is not so tolerant. (Hájek, 2011, p. 1)

And as we will see, nor is it implausible that some of the rules corresponding to these axioms and theorems constrain things too much - that is, the Kolmogorov axioms may not be the best explication of our pre-theoretical notion of probability.

7.3 Conditionalisation

Conditionalisation is the process of getting new probabilities from old ones by taking into consideration old *conditional* probabilities. Using principles of conditionalisation, we can model what happens to our subjective probabilities when we encounter new

[4] See e.g. Hájek (2012b, pp. 411-12).

[5] It may be thought that so-called 'preface paradoxes' undermine the idea that coherence is necessary for rationality/reasonableness. (I am grateful to J. Adam Carter for pointing this out to me.) Suppose Jo is an experienced and rigorous researcher. Experience has taught her that even well-researched books are highly likely to include at least one falsehood. For this reason, she believes that at least one of the claims in her new book will be false, and she prefaces it with an admission. At the same time, having thoroughly researched each of the claims, Jo seems to have good reason to believe that they are all true. It seems, then, that Jo has good reason to believe two incompatible propositions. Therein lies the paradox. However, it does *not* follow that Jo has good reason to assign a non-zero credence to the conjunction of these propositions, nor does it follow that she has good reason to assign a greater-than-1 credence to their disjunction, so it seems to me that such paradoxes do not undermine the idea that coherence is necessary for rationality/reasonableness.

evidence. The general idea is as follows: when exposed to new evidence for a proposition, one's new credence in that proposition should align with one's old credence in that proposition, conditional on the new evidence.

In order to understand what is meant by this, we need to know what is meant by conditional probability. Here is an example. Suppose you pick a sweet at random from a jar of 100 sweets, exactly 50 of which are aniseed flavoured. Call the proposition that the sweet picked will be aniseed flavoured 'A'. In that case, $P(A) = 0.5$. Now suppose exactly 30 of the sweets are coloured, and the rest are transparent. Call the proposition that the sweet picked will be coloured 'C'. In that case, $P(C) = 0.3$. So far, so simple. Now suppose exactly 20 of the coloured sweets are aniseed flavoured. In that case, $P(A \wedge C) = 0.2$. And finally, the conditional probability of C given A (i.e. the probability of C conditional on A) is 0.4. In formal notation: $P(C|A) = 0.4$.

In less clear cases, there is a handy formula we can use to calculate conditional probabilities from unconditional probabilities:

Ratio Formula $P(C|A) = P(A \wedge C) \div P(A)$.

To see that this formula works, we can plug in the values above, in which case we get $P(C|A) = 0.2 \div 0.5 = 0.4$.

We can also use this example to understand the process of conditionalisation. First, we need to know the value of $P(C|\neg A)$. Since only 10 of the 100 sweets are coloured and *not* aniseed flavoured, $P(\neg A \wedge C) = 0.1$. And since exactly half of the sweets are not aniseed flavoured, $P(\neg A) = 0.5$. So, by the Ratio Formula, $P(C|\neg A) = 0.1 \div 0.5 = 0.2$. Now, suppose my subjective probabilities currently align with all the probabilities and conditional probabilities just mentioned. Call my subjective probability distribution at this time Cr_o, where the subscript stands for 'old.' Thus, $Cr_o(A) = 0.5$, $Cr_o(C) = 0.3$, $Cr_o(C|A) = 0.4$, and $Cr_o(C|\neg A) = 0.2$. Suppose I now encounter new evidence, A, and my subjective probability in A goes up to 1. For example, suppose I randomly pick a sweet with my eyes closed and then eat it without seeing whether it is coloured. Suppose I am also so confident in my ability to identify flavours that I come to be certain that the sweet is aniseed flavoured. In formal notation: $Cr_n(A) = 1$, where the subscript stands for 'new.' In that case, my new subjective probability in C should be 0.4, equal to my old subjective *conditional* probability in C given A. In formal notation: $Cr_n(C) = 0.4$. More generally, for any propositions, A and C, when $Cr_o(A) \neq 1$ and $Cr_n(A) = 1$, $Cr_n(C)$ ought to update according to the following principle:

Simple Conditionalisation $Cr_n(C) = Cr_o(C|A)$.

Unfortunately, Simple Conditionalisation has an obvious problem, which is that it is only applicable when $Cr_n(A) = 1$. We rarely (if ever) assign such a high credence to a proposition. And indeed, we plausibly ought not to do so. As Lewis writes, '[t]o do so is to dismiss the genuine possibility that one has mistaken the evidence.' (D. K. Lewis, 1986b, p. 585) For example, I plausibly ought not to assign a subjective probability of 1 to the proposition that the sweet is aniseed flavoured - I may be good at identifying flavours, but I am not *infallible*.

With this in mind, what is needed is a more widely applicable principle - one which takes into account one's doubts about the evidence itself. For this, we turn to the following principle, suggested by Jeffrey (1965):

Jeffrey Conditionalisation $\mathrm{Cr}_n(C) = (\mathrm{Cr}_n(A) \times \mathrm{Cr}_o(C|A)) + (\mathrm{Cr}_n(\neg A) \times \mathrm{Cr}_o(C|\neg A))$.

Let us see how Jeffrey Conditionalisation works in practice. Suppose we have a situation like that described in the example above, except that upon eating the sweet, I update my subjective probability distribution so that $\mathrm{Cr}_n(A) = 0.9$. Given rule 7, it ought to be that $\mathrm{Cr}_n(\neg A) = 0.1$. Assuming I follow this rule, Jeffrey Conditionalisation then prescribes that $\mathrm{Cr}_n(C) = (0.9 \times 0.4) + (0.1 \times 0.2) = 0.36 + 0.02 = 0.38$. The principle takes into account not only my old conditional credence in C given A, but also my old conditional credence in C given *not-A*, and since the latter is slightly lower than the former, my new credence in C is slightly lower than 0.4.

When $\mathrm{Cr}_n(A) = 1$, Jeffrey Conditionalisation prescribes exactly the same new credence in C as Simple Conditionalisation does (assuming $\mathrm{Cr}_n(\neg A) = 0$, as per rule 7). Hence, Jeffrey Conditionalisation is an extension of Simple Conditionalisation: taken in conjunction with the rules of probability logic, it prescribes all the same updates in all the same situations *plus* some other updates in other situations. Of course, we rarely if ever know precisely what subjective probability we assign to pieces of evidence upon encountering them directly, and we rarely if ever know precisely what the relevant *conditional* probabilities are either.[6] Nonetheless, Jeffrey Conditionalisation gives us a plausible way of modelling what *should* be going on whenever we encounter new evidence.

7.4 Objective Probability

An objective probability function is one that represents the world itself as opposed to a subject's view of the world. The particular type of objective probability in which I couch the connection between conditionals and conditional probability is objective *chance*. Additionally, I endorse a Lewisian conception of objective chance as rational subjective probability conditional on history plus history-to-chance conditionals. In this section, I start to explain that Lewisian conception of objective chance.

[6] This is assuming that there are precise subjective probabilities to be known. Of course, one might disagree with this assumption.

One thing to note immediately is that, according to Lewis, objective chance may be distinguished from relative frequency, because we may sensibly talk of the (non-extremal) chances of types of events that never actually occur. In this vein, Lewis writes:

> [C]onsider unobtainium^{349}... there is not one atom of it in all of space and time. Its frequency of decay in a given time is undefined: 0/0. If there's any truth about its chance of decay, this undefined frequency cannot be the truthmaker. (D. K. Lewis, 1994, p. 477)

Unobtainium is (of course) not a real element, but our concept of chance is hardly contingent on its fictional status.

Lewis's conception of objective chance is underpinned by a principle he calls 'the Principal Principle,' initially presented in his (1980). The basic idea is this:

> [I]f a rational believer knew that the chance of decay was 50%, then almost no matter what else he might or might not know as well, he would believe to degree 50% that decay was going to occur. *Almost* no matter; because if he had reliable news from the future about whether decay would occur, then of course that news would legitimately affect his credence. (D. K. Lewis, 1994, p. 475)

The most important claim here is that it is rational for one's credence in a proposition to be equal to the chance one knows that proposition to have. This is absolutely fundamental to our understanding of chance. Given this, we can identify an objective chance function as one that is equivalent to the credence function of a rational agent with the right kind of knowledge. If somehow it were possible to know true propositions about the future that shed light on the outcomes of chancy processes, such knowledge would be *inadmissible* in the sense that it would "unfairly" affect a rational agent's credence in the relevant proposition. Thus, only knowledge pertaining solely to the past and/or present (i.e. the history of the world so far) is admissible. However, even if one were to have complete knowledge of the history of the world so far, that would still be insufficient for knowing the chances of events yet to come. Crucially, in order to have knowledge of such chances, one needs to know the complete theory of chance. As Lewis conceives it, this theory consists of a set of conditionals that license inferences from particular histories to particular chance-values for particular propositions.[7]

[7] This appeal to the complete theory of chance may have an air of circularity about it, but it is important to note that the Principal Principle is not intended as an *analysis* of chance; rather, it is merely intended as a specification of its identity conditions. Lewis writes:

> [C]ould we possibly get any independent grasp on this concept, otherwise than by way of the concept of chance itself?... [M]y provisional answer is: most likely not... (D. K. Lewis, 1980, p. 289)

Call the complete theory of chance 'T' and the complete history of the world 'H'. We can now express the Principal Principle more formally:

Principal Principle $\mathrm{Ch}(C) = \mathrm{Cr}(C|H \wedge T)$.

Here, Ch is an objective chance function, C is any proposition whatsoever, and Cr is a 'reasonable initial credence function.' (D. K. Lewis, 1980, p. 277) This latter phrase requires some explanation. An *initial* credence function is just one that is prior to any process of conditionalisation, i.e. the credence function of someone at the beginning of their credal journey. A *reasonable* initial credence function is one that is an epistemically good way to begin one's credal journey. More specifically, Lewis uses the term 'reasonable' here to imply *coherence*:

> A reasonable initial credence function is, among other things, a probability distribution... [that] obeys the laws of mathematical probability theory. (D. K. Lewis, 1980, p. 277)

As far as I can tell, Lewis uses 'reasonable' and 'rational' interchangeably in this context. And as we noted above, coherence may not be sufficient for rationality. In the next section, we look at an additional condition that many philosophers, including Lewis, have argued for as a constraint on rationality for subjective probability functions.

7.5 Regularity

A *regular* credence function obeys the following rules:

rule 9 If $\mathrm{P}(A) = 1$, then A is (metaphysically) necessary.

rule 10 If $\mathrm{P}(A) = 0$, then A is (metaphysically) impossible.

Each of these rules can be derived from the other together with rule 7, since the negation of a necessary proposition is an impossible proposition (and vice versa). They are the converses of rules 2 and 4 respectively, but they cannot be derived from any of the rules corresponding to the Kolmogorov axioms and theorems. Since rules 9 and 10 can be derived from each other together with rule 7, regularity is sometimes expressed either as rule 9 *or* as rule 10. Alternatively, it is sometimes expressed as the contrapositive of one of these rules, e.g. 'If A is metaphysically possible, then $\mathrm{P}(A) > 0$.' (Hájek, 2012b, p. 413) [I have changed his 'C' to 'P' and his 'X' to 'A', and removed his italics]. In this context, we will think of regularity as involving both rule 9 and rule 10, since rule 7 will not be questioned, and we will treat the contrapositives of these rules as equivalent formulations. Finally, as with the rules stated earlier, rules 9 and 10 have been expressed here in terms of metaphysical possibility. This is perfectly consistent with Lewis's formulation of the constraint: that the probability of a proposition is zero 'only if [it] is the empty proposition, true at no worlds...' (1980, p. 267) It is also consistent with Hájek's preferred formulation (quoted above), which is explicitly in terms of metaphysical possibility.

Despite the fact that they cannot be derived from the rules of logic that correspond to the Kolmogorov axioms, rules 9 and 10 are quite intuitive. As a constraint on rational credence, regularity has been endorsed by Lewis (1980), as well as by e.g. Shimony (1955) and Skyrms (1995).[8] Whereas a function that conforms to the rules of logic corresponding to the Kolmogorov axioms is said to be 'coherent', a function that conforms also to rules 9 and 10 is said to be 'strictly coherent' - see e.g. Hájek (2012b, p. 412).

There is some reason to be wary of regularity as a constraint on rational subjective probability. As Williamson writes:

> Regularity runs into notorious trouble when the set of possibilities is infinite, given the standard mathematics of probabilities, on which they are real numbers between 0 and 1... [S]uppose that a rotating pointer can stop at any point on a circle. As space is usually conceived, the circle comprises uncountably many points. For each point, it is neither objectively nor subjectively certain that the pointer will not stop at it. Yet on any real-valued probability distribution, for almost every point on the circle, the real-valued probability that the pointer will stop at it is 0. (Williamson, 2007, p. 173)

To accommodate cases like this, Lewis appeals to infinitesimal Cr-values, 'each infinitely close but not equal to zero'. (1980, p. 268) Conditional credences can then be defined even in cases like the one described by Williamson, where A is one of uncountably many possibilities. However, to allow infinitesimals as values of a probability function is to reject both the non-negativity axiom (which says that probabilities must be real numbers) and the additivity axiom:

> Kolmogorov's axiomatization stipulates that probability measures are real-valued; hence, according to that axiomatization, there can be no infinitesimal probabilities. We may wish to relax the stipulation... However, the stipulation is not idle... Kolmogorov also required probabilities to be *countably additive*. But the notion of countable additivity does not even make sense in

[8] Lewis's endorsement of regularity comes with a caveat. He thinks that regularity makes sense as a constraint on rationality 'if we are talking of everyday, less-than-ideal rationality', but he thinks it cannot be a constraint on *ideal* rationality. (1986b, pp. 586-587) According to Lewis, ideal rationality implies that one never mistakes the evidence, and is always certain that the evidence is true. Such an agent would come to have an irregular credence function because they would therefore assign a credence of 1 to contingent propositions.

> the context of non-standard analysis, and there is no obvious analogue of it. (Hájek, 2003, p. 280)

In other words, the introduction of infinitesimal credence-values is not to be taken lightly.

However, there are also compelling arguments to incorporate rules 9 and 10 into our probability logic. One such argument is as follows. Suppose rational credence functions are irregular, meaning they conform to neither rule 9 nor rule 10. In that case, $\text{Cr}(A) = 0$ for some rational credence function, Cr, and some metaphysically possible proposition, A. Now, if we take conditional probability to be defined by the Ratio Formula, so that $\text{Cr}(C|A) =^{\textbf{df}} \text{Cr}(A \wedge C) \div \text{Cr}(A)$, then $\text{Cr}(C|A)$ is undefined when $\text{Cr}(A) = 0$, since there is nothing that it means to divide by 0. This means one cannot update by conditionalising when $\text{Cr}(A) = 0$. But if A is metaphysically possible, then a rational agent ought to be able to conditionalise on it in cases where it turns out to be true. On the other hand, if rule 10 is a constraint on rational subjective probability, then no such problem cases arise, since evidence never comes in the form of metaphysical impossibilities. As Lewis writes:

> I should like to assume that it makes sense to conditionalize on any but the empty proposition. Therefore I require that Cr is *regular*: Cr(A) is zero, and Cr($C|A$) is undefined, only if A is the empty proposition, true at no worlds... The assumption that Cr is regular... is required as a condition of reasonableness: one who started out with an irregular credence function (and who then learned from experience by conditionalizing) would stubbornly refuse to believe some propositions no matter what the evidence in their favor. (D. K. Lewis, 1980, pp. 267-8) [I have changed his '*C*' to 'Cr', '*B*' to '*A*', and '*A*' to '*C*'.]

Since conditionalisation is highly plausible as a model of how we should update our credences, and since the Ratio Formula is a plausible definition of conditional probability, the argument goes that regularity must be a constraint on rational subjective probability.

This is a compelling argument, but in its current form, I do not wish to endorse it, since I do not agree that the Ratio Formula *defines* conditional probability. As Hájek (2003) points out, we have a pre-theoretical concept of conditional probability, and the Ratio Formula is just one attempt to explicate it. Pre-theoretically, one may reasonably think that the conditional probability of a proposition given itself is always 1, even if the unconditional probability of that proposition is 0. So the fact that the Ratio Formula leaves $\text{P}(C|A)$ undefined when $\text{P}(A) = 0$ is reason to doubt the Ratio Formula qua definition of conditional probability. This is not to doubt any of the solutions provided by the Ratio Formula - it is just to doubt that the Ratio Formula provides all the solutions available.

Nonetheless, a related argument can be made. Suppose we have a rotating pointer as per Williamson's thought experiment. Suppose a is one of infinitely many points on the circle at which the pointer (metaphysically) might stop, and suppose also that A is the proposition that the pointer will stop at a. Finally, suppose (for the sake of *reductio*) that $\mathrm{P}(A) = 0$; hence, $\mathrm{P}(\neg A) = 1$. By the Ratio Formula, $\mathrm{P}(A|\neg A) = \mathrm{P}(\neg A \wedge A) \div \mathrm{P}(\neg A) = 0 \div 1 = 0$. Hence, $\mathrm{P}(A|\neg A) = \mathrm{P}(A)$. In other words, the conditional probability of A given $\neg A$ is equal to the unconditional probability of A. But what this means is that A and $\neg A$ are probabilistically independent under P. And that, it seems to me, is absurd.

Admittedly, I do not think this is a *decisive* argument in favour of regularity as a constraint on rational credence. However, nor do I take any of the counterarguments against regularity to be decisive in the opposite direction. According to Lewis, rational credence is regular, and my point is just to show that there is nothing indefensible about this view. If we endorse Lewis's conception of rational credence *and* his conception of objective chance, then (by the Principal Principle) objective chance must also be regular, and as we will see, this allows us to draw a connection between conditionals and conditional probability.

Having said all this, Lewis later revises the Principal Principle due to concern regarding its compatibility with his view on the laws of nature. In the next section, I look at the revised version of the principle.

7.6 The Conditional Principle

As mentioned in section 3.6, Lewis endorses a Humean view of the laws of nature known as the best-system analysis, according to which something 'is a law iff it is a theorem of the best [deductive] system'. (1994, p. 478) Lewis's concern is that this view of laws does not accord with the Principal Principle in its original presentation because of a problem he calls 'undermining'. If the laws of nature - including probabilistic laws - are theorems of the best system, then they are determined in part by what happens in the future. But consider some counterfactual future, F, which results from some present or future chancy process, E, going differently. F will not happen (because it is counterfactual), but it currently has a non-zero chance of happening. The trouble is, if F is the type of future that affects the chances of outcomes of chancy processes like E, then were F to happen, its current chance would not be what it is. And since F results from E going differently, it is exactly the type of future that affects the chances of outcomes of chancy processes like E. As Lewis writes:

> '[S]uppose that we have a Humean analysis which says that present chances

> supervene upon the whole of history, future as well as present and past... Then different alternative future histories would determine different present chances... For instance, there is some minute present chance that far more tritium atoms will exist in the future than have existed hitherto, and each one of them will decay in only a few minutes. If this unlikely future came to pass, presumably it would complete a chancemaking pattern on which the half-life of tritium would be very much less than the actual 12.26 years... Although there is a certain chance that this future will come about, there is no chance that it will come about while still having the same present chance it actually has. (D. K. Lewis, 1994, p. 482)

On the one hand, F seems to be possible, because $\mathrm{Ch}(F) > 0$. On the other hand, F seems to be impossible, because if it were to occur, $\mathrm{Ch}(F)$ would not be the value that it is. Hence, on the best-system view of laws of nature, chances seem to undermine themselves.

Lewis's solution to the problem of undermining is as follows. As above, he identifies objective chance with the credence of a rational agent in possession of the right kind of knowledge, including knowledge of the present chances of individual propositions. This allows Lewis to give the following diagnosis of the problem of undermining: given that the chances are determined in part by the way the future actually pans out, an agent cannot know the present chances of individual propositions without possessing some of the wrong kind of knowledge. In other words, knowledge of present chances is *inadmissible*.

This diagnoses the problem of undermining, but it also makes the Principal Principle incompatible with the best-system analysis of natural laws. To resolve this tension, Lewis argues that the original version of the Principal Principle is not quite right, and he proposes a corrected version of the principle, which I will call the 'Conditional Principle'. This corrected version of the principle is as follows:

Conditional Principle $\mathrm{Ch}(C|T) = \mathrm{Cr}(C|H \wedge T)$.

This principle says that the *conditional* chance of a proposition, *given the complete theory of chance*, comes from any reasonable initial credence function by conditionalising on the complete history of the world up to the time, together with the complete theory of chance for the world. By conditionalising on the complete theory of chance on the left hand side as well as the right, the Conditional Principle removes the potential for counterfactual futures to undermine present chances - the present chances are conditional on the probabilistic laws of nature as determined by the *actual* future, so counterfactual futures simply cannot affect them.

While Lewis considers the Conditional Principle to be more accurate than the original Principal Principle, he still considers the original version to be our key to *understanding* objective chance. This is justified partly on the basis that he takes inadmissibility to admit of degrees. Note that the solution to the problem of undermining does not get rid of the inadmissibility of the complete theory of chance - the Conditional Principle guards against the problem of undermining, but undermining was only the symptom, not the disease. In fact, Lewis does not think the complete theory of chance can be made to be perfectly admissible; rather, he thinks inadmissibility admits of degrees, and the extent to which T is inadmissible is negligible. He writes:

> [N]ear-admissibility may be good enough. If T specifies that the present chance of C is Ch(C), and if T is nearly admissible relative to C, then the [Principal Principle] will hold, if not exactly, at least to a very good approximation. If information about present chances is never perfectly admissible, then the Principal Principle never can apply strictly. But the Principle applied loosely will very often come very close, so our ordinary reasoning about chance and credence will be unimpaired. (D. K. Lewis, 1994, p. 486) [I have changed his 'E' to 'T', 'A' to 'C', 'P' to 'Ch', and 'C' to 'Cr'.]

Let us take stock. Objective chance, according to Lewis, comes from any reasonable initial credence function by conditionalising on the right kind of knowledge. The right kind of knowledge seems to be knowledge of the history of the world plus knowledge of history-to-chance conditionals comprising the complete theory of chance. But given a best-system analysis of natural laws, the complete theory of chance is inadmissible, because it is determined in part by the chancemaking patterns of the future. This gives rise to the problem of undermining futures: taking the complete theory of chance into account means that counterfactual futures are rendered both (metaphysically) possible *and* (metaphysically) impossible by their present chances. The solution to the problem, according to Lewis, is to conditionalise chances on the theory of chance as determined by the *actual* future, thereby protecting the present chances from counterfactual futures that undermine them. However, this does not render the original Principal Principle obsolete, since it may still very closely approximate the right answer.

For what it is worth, I do not think Lewis needs to defend this conception of chance via the Principal Principle in particular. There has long been a tendency to treat unconditional probabilities as more fundamental than conditional probabilities, as demonstrated by the fact that the Ratio Formula has come to be seen as a definition or analysis of conditional probability.[9] However, I think this is exactly the wrong way round, and if we reverse the idea that unconditional probabilities are more fundamental than conditional probabilities, then the Conditional Principle starts to look less like an unfortunate

[9] For example, after describing the reasonable initial credence function that features in the Principal Principle, Lewis (1980, p. 267) writes that 'the corresponding conditional credence function is defined simply as a quotient of unconditional credences'.

complication and more like a genuine insight into the nature of chance.[10] Indeed, if conditional probabilities are the more fundamental, then a principle connecting conditional credence with conditional chance is exactly what we ought to hope for. In chapter 9, I use this idea to clarify a recent philosophical debate on the nature of risk.

At any rate, I think Lewis is right to view objective chance (conditional or unconditional) as rational credence conditional on the right kind of knowledge. In what follows, I take this Lewisian conception of chance for granted.

7.7 The Connection

In this section, I show that the unified strict analysis of conditionals that I defend (which says that $A > C$ is true iff $A \wedge \neg C$ is metaphysically impossible) allows us to draw a connection between conditionals and conditional chances if chance is conceived as per Lewis's Conditional Principle. That connection (which I have called 'McCardel's Thesis') is that $A > C$ is true iff $\mathrm{Ch}(C|A) = 1$.

To demonstrate this, I will have to appeal to an uncontroversial rule of probability logic governing conditional probabilities in particular:

rule 11 $\mathrm{P}(C|A) + \mathrm{P}(\neg C|A) = 1$.

This is the *complement* rule for conditional probabilities, and it says that the conditional probability of C given A plus the conditional probability of $\neg C$ given A equals 1. I will also have to appeal to rules 4 and 10, the latter being the regularity constraint.

First, let us consider the claim that $\mathrm{Ch}(C|A) = 1$ *if* $A \wedge \neg C$ is metaphysically impossible:

Suppose $A \wedge \neg C$ is metaphysically impossible.

By rule 4, if $A \wedge \neg C$ is metaphysically impossible, then $\mathrm{Ch}(A \wedge \neg C) = 0$.

Hence, $\mathrm{Ch}(A \wedge \neg C) = 0$.

By the Ratio Formula, $\mathrm{Ch}(\neg C|A) = \mathrm{Ch}(A \wedge \neg C) \div \mathrm{Ch}(A)$.

Hence, $\mathrm{Ch}(\neg C|A) = 0 \div \mathrm{Ch}(A)$.

Hence, $\mathrm{Ch}(\neg C|A) = 0$.

By rule 11, $\mathrm{Ch}(C|A) + \mathrm{Ch}(\neg C|A) = 1$.

Hence, $\mathrm{Ch}(C|A) = 1$.

Therefore, $\mathrm{Ch}(C|A) = 1$ *if* $A \wedge \neg C$ is metaphysically impossible.

Now let us consider the claim that $\mathrm{Ch}(C|A) = 1$ *only if* $A \wedge \neg C$ is metaphysically impossible:

[10] Popper (1959) famously axiomatises a view of conditional probability by defining functions from *pairs* of propositions to numbers. These functions have come to be known as 'Popper functions', and the view can be seen as capturing the idea that conditional probabilities are fundamental.

Suppose Ch($C|A$) = 1.

By rule 11, Ch($C|A$) + Ch($\neg C|A$) = 1.

Hence, Ch($\neg C|A$) = 0.

By the Ratio Formula, Ch($\neg C|A$) = Ch($A \wedge \neg C$) ÷ Ch(A).

Hence, Ch($A \wedge \neg C$) ÷ Ch(A) = 0.

Hence, Ch($A \wedge \neg C$) = 0.

By rule 10, if Ch($A \wedge \neg C$) = 0, then $A \wedge \neg C$ is metaphysically impossible.

Hence, $A \wedge \neg C$ is metaphysically impossible.

Therefore, Ch($C|A$) = 1 *only if* $A \wedge \neg C$ is metaphysically impossible.

On the supposition that Ch is regular, we have established the claim that Ch($C|A$) = 1 *if and only if* $A \wedge \neg C$ is metaphysically impossible. Given the unified, metaphysically strict analysis of conditionals that I defend, this means that Ch($C|A$) = 1 iff $A > C$ is true. This is good news for the analysis of conditionals that I defend.

7.8 The Triviality Results

Perhaps the main threat to the connection established in the previous section is the possibility that there may be a counterargument to it based on a triviality result, just as there have been many such counterarguments against Stalnaker's Thesis. In this section, I explore that possibility.

In formal notation, Stalnaker's Thesis is as follows:

Stalnaker's Thesis P($A > C$) = P($C|A$), except where P(A) = 0.

On the face of it, it looks highly plausible, and also quite similar to the connection established in the previous section. Unfortunately, Lewis (1976) presents a strong *reductio* against Stalnaker's Thesis. Given only a few fairly uncontroversial assumptions, Lewis uses Stalnaker's Thesis to reach the following conclusion: for any coherent probability function, P, and any propositions, A and C, if P($A \wedge C$) and P($A \wedge \neg C$) are both positive, then P($C|A$) = P(C); in other words, A and C are probabilistically independent under P. But, as Lewis observes, '[t]hat is absurd.' (1976, p. 300) Here is a straightforward example to demonstrate its absurdity. Suppose I pick a card at random from a normal deck. Let A be the proposition that the card is red, and let C be the proposition that the card is a diamond. In that case, Ch($A \wedge C$) and Ch($A \wedge \neg C$) are both positive, but Ch($C|A$) ≠ Ch(C), since the truth of A affects the chance of C. In other words, A and C are *not* probabilistically independent with respect to Ch (a coherent probability function), and it would be absurd to endorse a thesis that implies otherwise.

More generally, Lewis points out that one can construct a counterexample to the conclusion using any language capable of expressing three possible but pairwise incompatible propositions. (Propositions are pairwise incompatible iff no two of them can be true at the same time, for example: $P \wedge Q$, $\neg P \wedge Q$, $P \wedge \neg Q$, and $\neg P \wedge \neg Q$.) Suppose B, C, and D are possible but pairwise incompatible, and suppose $A \equiv B \vee C$. From the fact that B, C, and D are possible but pairwise incompatible, we can infer that none of them are negations of each other, and that none of them are necessary truths. From these facts, we can also infer that A is not a necessary truth. Hence, there are coherent probability functions according to which $\mathrm{P}(A \wedge C)$ and $\mathrm{P}(A \wedge \neg C)$ are both positive, but $\mathrm{P}(C|A) \neq \mathrm{P}(C)$. See Lewis (1976, p. 300) for more details.

Lewis's *reductio* gives us reason to be wary of Stalnaker's Thesis. More specifically, it gives us reason to believe that Stalnaker's Thesis cannot be true for all coherent probability functions - at least, not if we interpret it as a thesis about the conditionals of *our* language. If the argument above is correct, then Stalnaker's Thesis can only be true for all probability functions if we interpret it as a thesis about the conditionals of a *trivial* language - one in which it is impossible to express three possible but pairwise incompatible propositions.

One might still wonder whether Stalnaker's Thesis can be true (of the conditionals in *our* language) for some proper subset of all probability functions, e.g. rational subjective probability functions. Unfortunately, however, the *reductio* works just the same for any set of coherent probability functions that is *closed under conditionalising* - i.e. any set of coherent functions such that conditionalising from one member takes you to another member in the set. In fact, Lewis shows that if Stalnaker's Thesis is true for some proper subset of all coherent probability functions closed under conditionalising, then not only must we interpret it as a thesis about the conditionals of a trivial language, every function in the subset must also be trivial in the sense that it has at most four different values.[11]

The *reductio* born of these triviality results has been highly persuasive - even Stalnaker has given up Stalnaker's Thesis. Of course, no argument is indisputable. One might question some of the assumptions used by Lewis to reach the absurd conclusion, for example.[12] Alternatively, one might think there is some relevant subset of probability functions that is not closed under conditionalising. One might even accept that the only relevant probability functions are trivial ones according to which all probabilities

[11] To see this, note that if some function in the subset were to have more than four values, then at least two of its values, x and y, would be such that $0 < x < 1$, $0 < y < 1$, and $x + y \neq 1$. But suppose $\mathrm{P}(E) = x$ and $\mathrm{P}(F) = y$ for some E and F. In that case, at least three of $\mathrm{P}(E \wedge F)$, $\mathrm{P}(\neg E \wedge F)$, $\mathrm{P}(E \wedge \neg F)$, and $\mathrm{P}(\neg E \wedge \neg F)$ would be positive; hence, at least three of these (pairwise incompatible) conjunctions would be *possible*. But that contradicts the fact that the language is trivial. See Lewis (1976, p. 302).

[12] Lepage (2015) points out that Lewis assumes the 'expansion by cases' principle, which says that $A \equiv (A \wedge B) \vee (A \wedge \neg B)$. As Lepage shows, this principle is invalid in intuitionistic logic.

are either 0 or 1, in which case it cannot be that both $P(A \wedge C)$ and $P(A \wedge \neg C)$ are positive. Clearly, however, the triviality results are not to be dismissed lightly, and the view of conditionals that I defend in this thesis is not so far removed from Stalnaker's Thesis that I need not worry about the triviality results at all.

Let us compare my view with Stalnaker's Thesis. Whereas Stalnaker claims that $P(A > C) = P(C|A)$, I claim that $A > C$ is true iff $P(C|A) = 1$. These claims are closely related, but precisely *how* closely related depends on one's background suppositions. Suppose propositions pertaining solely to the past and/or present have objective probabilities of either 1 or 0 depending on whether they are true or false respectively.[13] In support of this supposition, consider e.g. the following passage of Lewis's:

> We ordinarily think of chance as time-dependent... Suppose you enter a labyrinth at 11:00 a.m., planning to choose your turn whenever you come to a branch point by tossing a coin. When you enter at 11:00, you may have a 42% chance of reaching the center by noon... At 11:49 you reach the center; then and forevermore your chance of reaching it by noon is 100%. (D. K. Lewis, 1980, p. 91)

Suppose also that propositions pertaining partly to the future are neither true nor false, but only more or less (objectively) probable. Neither of these suppositions is absurd, but together, they imply that a proposition (conditional or unconditional) is true iff it has a probability of 1. In that case, Stalnaker's Thesis *implies* my thesis. Put simply: suppose $A > C$ is true iff $P(A > C) = 1$; in that case, if $P(A > C) = P(C|A)$, then $A > C$ is true iff $P(C|A) = 1$.

What would this mean for McCardel's Thesis? By itself, nothing bad at all. What would be bad is if my thesis implied Stalnaker's, for then it would be susceptible to all the same counterarguments. Indeed, the fact that my view is implied by Stalnaker's under these suppositions actually helps to explain what it is about Stalnaker's Thesis that *seems right*. It may even help to explain what it is about Adams' Thesis that seems right, since Stalnaker's Thesis can be seen as motivating Adams' Thesis.

Are there any suppositions under which McCardel's Thesis implies Stalnaker's Thesis? Yes. On the supposition that all propositions are either true or false (contra the other supposition above), my view implies that a conditional is false iff $P(C|A) \neq 1$. And if, in addition to this, P has only two values, 0 and 1, then my view is equivalent to Stalnaker's. But in that case, my view is not susceptible to Lewis's *reductio*, because we are simply accepting that the probability function in question is a trivial one.

One last point before moving on: if the unified, metaphysically strict analysis that

[13] Given regularity, this supposition makes any conditional with a false antecedent about the past *vacuously* true (because its antecedent is metaphysically impossible), and any conditional with a true consequent about the past *trivially* true (because its consequent is metaphysically necessary).

I defend is right, then it should be no surprise that McCardel's Thesis and Stalnaker's Thesis have different consequences. When A and C are both contingent, $\mathrm{Ch}(C|A)$ may be anything in between 0 and 1, whereas the chance of $\Box(A \supset C)$ ought always to be either 0 or 1. On the conception of chance outlined above, it makes no sense to expect a non-extremal answer to the question 'What is the chance that $A \supset C$ is a metaphysical necessity?' Perhaps one can rationally assign non-extremal subjective probabilities to metaphysical modal propositions under some circumstances, but not if one knows all of history and the complete theory of chance. If we conceive of objective chance in a Lewisian way, then metaphysical modal propositions like $\Box(A \supset C)$ have a chance of either 0 or 1.

7.9 Edgington's Objections

In her (1995) *Mind* article 'On Conditionals', Edgington points out that some philosophers have suggested that a conditional is 'true iff the objective probability of [C] given A is sufficiently high.' (1995, p. 292) She attributes the view to Blackburn (1986, pp. 213-5) and Woods (1997), and says also that it 'crops up orally from time to time'. (1995, p. 292, footnote 55) For this reason, let us call it 'Common Thesis', defining it as follows:

Common Thesis $A \to C$ is true iff $\mathrm{P}(C|A)$ is sufficiently high.

She then goes on to argue that such a view is objectionable. This should give us pause, since McCardel's Thesis implies a version of Common Thesis where 'sufficiently high' is taken to mean '1'.

One of the issues that Edgington has with Common Thesis is that it conflicts with a principle she calls 'The Thesis' (1995, p. 263), but which for clarity's sake I will call:

Edgington's Thesis $\mathrm{Cr}(A \to C) = \mathrm{Cr}(A \wedge C) \div \mathrm{Cr}(A)$.

The reason it conflicts with Edgington's Thesis is as follows. Suppose I am certain that the objective probability of A is greater than zero, and the objective conditional probability of C given A is some number between 0 and 1, say 0.9. Suppose I am also certain that 0.9 is *in*sufficiently high for the truth of $A \to C$. In that case, my credence in $A \to C$ should be 0. But by Edgington's Thesis and the Ratio Formula, it should be 0.9.

Why believe Edgington's Thesis? As should be clear, Edgington's Thesis can be derived from the Ratio Formula together with Stalnaker's Thesis. Of course, we have just seen that Stalnaker's Thesis is problematic, so this route to Edgington's Thesis is also problematic. Still, Edgington's Thesis is intuitive by itself, and according to Edgington it can be derived without recourse to Stalnaker's Thesis. For example, we can supposedly derive it from Ramsey, who declares that 'Degree of belief in (p and q) =

degree of belief in p × degree of belief in q given p.' (Ramsey, 1931b, p. 181) Edgington takes 'q given p' to be a 'mere stylistic variation' (1995, p. 262) on 'q if p', in which case Edgington's Thesis can be derived from Ramsey's claim.

However, on the metaphysically strict analysis, 'q given p' is *not* a 'mere stylistic variation' on 'q if p': 'if' conveys a metaphysical connection that 'given' does not. Hence, this particular route to motivate Edgington's Thesis does not work on the view that I have been defending. In the absence of a compelling reason to accept Edgington's Thesis beyond intuitive appeal, I am quite happy to reject it.

The other issue that Edgington has with Common Thesis is that it

> has the consequence that the truth of A&[C] is compatible with the certain falsity of "If A, [C]". Not everyone minds that. *I* think it's wrong for me to say "It is certainly false that, if you approach, the dog will bite", when I know that the objective conditional probability of its biting, given that you approach, is 0.5; and further, to admit no error when you approach and are bitten—to stick to my judgement that the conditional was certainly false. But not everyone agrees with me. (Edgington, 1995, p. 293)

This is a reasonable objection. Nonetheless, I am committed to the belief that the conditional 'If you approach, the dog will bite' is false in this scenario. In fact, I think its falsehood is even quite intuitive when framed in the right way. Given that I know the relevant objective conditional probability is 0.5, we are dealing with a genuinely chancy system. But in that case, there is no fact of the matter about what *would* happen if one *were* to approach. As Hájek says: 'to say that there is a fact of the matter of how the toss would land is to deny that the coin is a chancy system.' (2014, p. 7) This undermines the 'would'-conditional 'If you were to approach, then the dog would bite'. But given a unified analysis, that means the indicative conditional is undermined as well.

I think the trouble with the statement made in the thought experiment is simply that it is misleading. First of all, it is ripe for misinterpretation, because you (the hearer) are likely to confuse it with the conditional 'If A, then not-C', since people are prone to exactly this sort of scope fallacy. Secondly, the assertion flouts Gricean norms of conversational implicature, because I know something that is more relevant and more informative: namely, that the objective conditional probability of the dog's biting, given that you approach, is 0.5. I should have asserted *that* instead.

7.10 Conclusion

In this chapter, I began by explaining some plausible rules of probability logic and other key concepts in the theory of probability. This paved the way for explaining Lewis's conception of objective chance, and the regularity constraint that comes with it. I have shown that, on this conception of chance, if the strict analysis of conditionals that I defend is true, then we can draw a connection between conditionals and conditional probability such that $A > C$ is true iff $\mathrm{P}(C|A) = 1$. Finally, I hope to have warded off concerns based on the triviality results, and to have shown that this connection can be defended even in the face of compelling objections made by Edgington. In the next chapter, I use this connection to supplement a view of causal explanation presented by Jackson and Pettit in their influential (1990) paper ‘Program Explanation: A General Perspective’.

Chapter 8

Program Explanation and Modal Information

Conditionals have played a prominent role in theories relating to causation (e.g. Lewis (1973)) and causal explanation (e.g. Woodward (2003)). One influential view of causal explanation in which they do *not* currently feature very prominently is the view expressed in Jackson and Pettit's 'Program Explanation: A General Perspective' (1990). Therein, the authors offer a novel perspective on program[1] explanation: a type of explanation that involves appealing to causally non-efficacious properties that nonetheless make 'suitably probable' the instantiation of causally efficacious ones. (Jackson & Pettit, 1990, p. 114) This type of explanation is useful, the authors claim, because it conveys modal information that might otherwise not be conveyed.

In this chapter, I show that the view of conditionals defended throughout this thesis can supplement Jackson and Pettit's view. In particular, I use McCardel's Thesis (see chapter 7) to infer the truth of a related conditional in cases of program explanation where the non-efficacious properties make the instantiation of the relevant efficacious properties *maximally* probable. Analysing that conditional as per the metaphysically strict analysis offers us a novel perspective on Jackson and Pettit's novel perspective. In particular, it allows us to succinctly explain *why* program explanation conveys the modal information it conveys.

I begin by explaining the way that Jackson and Pettit conceive of *causal efficacy* and *causal explanation*. Careful consideration of these ideas leads to a paradox, the upshot of which is that our ordinary causal explanations appear not to be causal explanations at all. In section 8.2, I explain two versions of this paradox, the second of which is tailored to withstand a solution to the first. I then explain Jackson and Pettit's solution to both versions of the paradox, which in turn helps us to understand their novel perspective on program explanation as a way of conveying modal information. In section 8.4, I explain

[1] For the sake of consistency with Jackson and Pettit, I will continue to use this spelling of the word.

my novel perspective on *their* novel perspective: that the truth of a related conditional succinctly explains why this modal information is conveyed. Finally, I conclude that this helps us to make sense of Jackson and Pettit's view, which in turn helps us to justify our ordinary practice of causal explanation.

8.1 Causal Efficacy and Causal Explanation

To understand the paradox presented by Jackson and Pettit, we first need to understand what they mean by 'causal efficacy'. We tend to conceive of the physical world as a series of causal interactions between things. Let us say, metaphorically, that each of these things is a link on a causal chain. To be causally efficacious with regard to some effect is to feature on the anterior part of the causal chain: the part that precedes the effect. For example, if a pool player pots the blue ball with the white ball, then the white ball, the cue, and the pool player all feature as links on the anterior part of the causal chain. Importantly, Jackson and Pettit presume that it is the *properties* of these objects (as opposed to the objects themselves) that are causally efficacious. For example, the blue ball moved because of the *momentum* of the white ball that struck it. Similarly: a swimmer clenched because of the *temperature* of the water; a driver accelerated because of the *colour* of the traffic light; and so on. In sum, according to Jackson and Pettit, to be causally efficacious with regard to an effect is to feature as a property of a link on the part of the causal chain that precedes the effect.

We also need to have at least a rough understanding of what Jackson and Pettit mean by 'causal explanation.' Jackson and Pettit make the uncontroversial assumption that a causal explanation of an effect must tell us about something of causal *relevance* to the effect. As they point out, one might reasonably assume that being causally efficacious is the *only* way to be causally relevant with regard to an effect. (1990, p. 111) However, we need not rely on this assumption in order to explicate their conception of causal explanation. Instead, let us continue to use the causal chain metaphor.

Suppose we are trying to find out how some effect came to be. What we need is some information about the part of the causal chain that precedes the effect. On a simple but common view, causal explanation regarding an effect is just the provision of this kind of information. Such information may tell us about the links that can be found on the anterior part of the chain, and the properties they instantiate; alternatively, it may also tell us about the links or properties that *cannot* be found on the anterior part of the chain. For example, either 'I struck the white ball' or 'I didn't strike the white ball' may count as a true causal explanation of potting the blue ball. Discovering what cannot be found on the chain is useful insofar as it dispels erroneous expectations. 'I didn't strike the white ball,' for example, may dispel the erroneous expectation that one's shot did not violate any rules. However, such causal explanations do not tell us how the effect came to be. On Jackson and Pettit's view, a *complete* causal explanation (i.e. one that describes a complete causal chain up to the effect) must tell us what *can* be found. More formally, it must tell us that an effect e was preceded by some particular thing c_1, instantiating some particular property F_1, and that c_1 was preceded by c_2, instantiating another particular property F_2, and so on, either to a first cause, or *ad infinitum*.

Of course, our everyday causal explanations are never complete in this sense. Rather, they aim at levels of completeness that are appropriate for their context. Causal explanation, like any form of communication, is governed by Grice's second maxim of Quantity: 'Do not make your contribution more informative than is required.' (1967b, p. 26) Given our characterisation of a complete causal explanation, one way in which a causal explanation may be incomplete is if it does not specify all the links and all the properties on the anterior part of the causal chain. But according to Jackson and Pettit, another way in which a causal explanation may be incomplete is if it fails to make reference to the right kind of property. Suppose we have a property F such that if F is causally efficacious, then another property G is (also) causally efficacious. Suppose that F neither precedes G on the causal chain nor combines with G in order to achieve causal efficacy. According to Jackson and Pettit, any such property F is not really causally efficacious. (1990, p. 108)

To see why Jackson and Pettit make this claim, let us consider an example. Consider again the claim that a driver may accelerate because of the *colour* of the traffic light. If the colour property is causally efficacious with regard to the driver's acceleration, then the properties of the fundamental particles that constitute the traffic light are also causally efficacious with regard to the driver's acceleration. Take one such fundamental property; the colour of the traffic light neither precedes the fundamental property on the causal chain nor combines with it in order to achieve causal efficacy. The colour property thus seems to be redundant in a significant way: a causal explanation of the driver's acceleration may be complete even if it does not make reference to the colour property of the traffic light. Jackson and Pettit ascribe causal efficacy only to those properties that cannot be omitted from complete causal explanations. Following Jackson and Pettit, let us call the colour of the traffic light a 'higher-order' property, and the properties of the traffic light's fundamental particles 'lower-order' properties. (1990, *passim*) On Jackson and Pettit's view, higher-order properties are not causally efficacious; only lower-order properties are. (1990, p. 108)

This presents us with a paradox. Before stating it in full, let us recap what has been said so far. Initially, it was said that, on Jackson and Pettit's view, to be causally efficacious with regard to an effect is to feature as a property of a link on the part of the causal chain that precedes the effect. It was then said that a causal explanation regarding an effect must tell us about something of causal relevance to the effect, and that one might reasonably assume that the only way to be causally relevant is to be causally efficacious. Lastly, it was said that, on Jackson and Pettit's view, only lower-order properties are genuinely causally efficacious.

8.2 A Paradox of Causal Explanation

The above leads us to expect that causal explanations must provide us with information about lower-order properties. And yet, ordinary causal explanations rarely (if ever) do so - most of the time, they make reference to properties that are higher-order, such as the colour of the traffic light, the temperature of the water, and so on. This makes it difficult to see how ordinary causal explanations qualify as causal explanations. Premise by premise, the paradox can be stated as follows:

(I) *Something is a causal explanation of an effect only if it refers to a property that is causally relevant with regard to the effect.*

(II) *A property is causally relevant with regard to an effect only if it is causally efficacious with regard to the effect.*

(III) *A property is causally efficacious only if it is lower-order.*

Therefore,

(IV) *Something is a causal explanation of an effect only if it refers to a property that is lower-order.*

But,

(V) *Most causal explanations do not refer to properties that are lower-order.*

If all of (I)-(III) are true, then making reference to a lower-order property is a necessary condition on being a causal explanation. But since this is inconsistent with (V), and since (V) is clearly true, it must be that at least one of the first three propositions is false.

One way to resolve the paradox is to argue that higher-order properties *are* causally efficacious, and that (III) is therefore false. However, Jackson and Pettit do not make this move, because they think that to attribute causal efficacy to higher-order properties such as colour is to needlessly proliferate causal efficacy. They also argue that, even if one is happy to proliferate causal efficacy, a version of the paradox persists. As above, lower-order properties play a distinct and important role in that they cannot be omitted from complete causal explanations. Jackson and Pettit take this as reason to think that even if higher-order properties are causally efficacious, lower-order properties exercise a distinct and more fundamental kind of causal efficacy. Call this 'primitive' causal efficacy. The paradox may then be reconstructed as follows:

(I) *Something is a causal explanation of an effect only if it refers to a property that is causally relevant with regard to the effect.*

(II)* *A property is causally relevant with regard to an effect only if it is primitively causally efficacious with regard to the effect.*

(III)* *A property is primitively causally efficacious only if it is lower-order.*

Therefore,

(IV) *Something is a causal explanation of an effect only if it refers to a property that is lower-order.*

But,

(V) *Most causal explanations do not refer to properties that are lower-order.*

This version of the paradox is able to withstand the solution to the original version of the paradox. One consequence of the adjustment, however, is that (II)* is less plausible than (II): being primitively causally efficacious with regard to an effect is not obviously the only way to be causally relevant with regard to the effect. This points us in the direction of a solution: Jackson and Pettit argue that there is a way for a property to be causally relevant besides being primitively causally efficacious. In fact, they argue that there is a way for a property to be causally relevant besides being either primitively *or* non-primitively causally efficacious. In other words, they reject both (II)* *and* (II), as I explain in the next section.

8.3 A Solution to the Paradox

According to Jackson and Pettit, another way in which a property can be causally relevant with regard to an effect is by programming for a property that is causally efficacious with regard to the effect. This requires some explanation. Let us say, along with Jackson and Pettit, that a property *F programs for* a property *G* iff *F* makes it 'suitably probable' (1990, p. 114) that some member of a set *X* to which *G* belongs is instantiated. For example, one can think of the general colour property *green* as programming for the more specific shade of green that is instantiated by the traffic light: the instantiation of green by the traffic light makes it suitably probable (indeed, ensures) that some more specific shade of green is instantiated. More importantly for our purposes, one can also think of green as programming for the lower-order properties of the fundamental particles that constitute the traffic light: the instantiation of green by the traffic light makes it suitably probable (indeed, ensures) that some particles in the area instantiate some lower-order properties belonging to a particular set.[2] This illustrates the more general point that higher-order properties program for lower-order properties.

Properties that program for other properties are called 'program properties', and explanations that appeal to such properties are called 'program explanations'. Jackson and Pettit first explained their view of program explanation in a previous paper called 'Functionalism and Broad Content' (1988). Their purpose in 'Program Explanation: A General Perspective' is not just to explain this view of program explanation, but (i) to present it as the solution to the above paradox, and (ii) to use the paradox to illustrate a novel perspective on the view. With regard to the first point: program explanation resolves the paradox by identifying a way in which a property may be causally relevant with regard to an effect without being causally efficacious with regard to the effect. The colour property *green*, for example, is causally relevant with regard to the driver's acceleration because it programs for some lower-order properties that are causally efficacious with regard to the driver's acceleration. Importantly, this helps to justify our ordinary practice of giving causal explanations in terms of higher-order properties like the colour of the traffic light, the temperature of the water, the momentum of the white ball, and so on.

With regard to the second point: by appealing to a property that programs for these lower-order properties, program explanation provides *modal* information about the part of the causal chain that precedes the effect. As Jackson and Pettit write:

> [T]o explain something is to provide information on its causal history... A program explanation provides a different sort of information from that which is supplied by the corresponding process account and therefore a sort of information which someone in possession of the process account may lack. The process story tells us about how the history actually went... A program account tells us information about how that history might have been. (Jackson & Pettit, 1990, p. 117)

According to Jackson and Pettit, program explanation thereby plays a distinct and important role, since causal explanation that makes reference solely to causally efficacious properties (what we might call 'process explanation') does *not* seem to provide modal information.

I think there is a succinct way to explain why program explanation conveys the modal information that it does. In the following section, I use the account of conditionals

[2] It is an unfortunate coincidence that *green* is being represented by the letter '*F*', whereas the *fundamental* properties are being represented by the letter '*G*'. I would flip these letters around, but that would cause confusion when comparing my explanation of Jackson and Pettit's view with their own explanation of their view - the letters I am using here match the letters that they use in 'Program Explanation: A General Perspective' (1990).

defended throughout this thesis to supplement Jackson and Pettit's view. In short, I offer a novel perspective on their novel perspective.

8.4 A Novel Perspective

Jackson and Pettit use the following example to further illustrate the kind of modal information provided by program explanation:

> The process story tells us about how the history actually went: say that such and such particular decaying atoms were responsible for the radiation. A program account tells us about how that history might have been... telling us for example that in any relevantly similar situation, as in the original situation itself, the fact that some atoms are decaying means that there will be a property realized... which is sufficient in the circumstances to produce radiation. In the actual world it was this, that and the other atom which decayed and led to the radiation but in possible worlds where their place is taken by other atoms, the radiation still occurs. (Jackson & Pettit, 1990, p. 117)

Here, Jackson and Pettit say that program explanation tells us about what 'might have been', but I think this is ripe for misinterpretation. When one comes to know a program explanation, one's set of live possibilities does not *expand* to include all those worlds consistent with the program explanation; rather, it *contracts* to include *only* those worlds consistent with the program explanation. After all, one does not begin by assuming that the actual causal history is the only possible causal history; rather, one begins by assuming that there are many possible causal histories. For this reason, the modal information conveyed by a program explanation might more naturally be captured in terms of what might *not* have been. And this is exactly the sort of modal information that is conveyed by a metaphysically strict conditional.

Consider some higher-order property F in terms of which an ordinary causal explanation may be given. Let G be a lower-order property such that G is causally efficacious with regard to the effect being explained if F is causally efficacious with regard to the effect being explained. According to Jackson and Pettit, F programs for G iff there is a suitably high conditional probability that a member of X is instantiated *given* that F is instantiated, where X is a set to which G belongs. Let $\mathbf{F}$ be the proposition that F is instantiated, and let $\mathbf{X}$ be the proposition that some member of X is instantiated; hence, F programs for G iff $\mathrm{P}(\mathbf{X}|\mathbf{F})$ is suitably high. Now suppose this conditional probability is not only suitably high, but *maximally* high. Given McCardel's Thesis (see chapter 7), the following conditional is true:

$\mathbf{F} \rightarrow \mathbf{X}$ *If F is instantiated, then some member of X is instantiated.*

On the metaphysically strict analysis of indicative conditionals, this is a modal proposition: it tells us that it is metaphysically impossible that *F* is instantiated and no member of *X* is instantiated. More relevantly, its truth prior to *F*'s instantiation tells us *after* the instantiation that it *could not have been* that *F* was instantiated and no member of *X* was instantiated. This, it seems to me, is the core of the modal information conveyed by a program explanation in terms of *F*.

Some readers may object to my interpretation of the modal information conveyed by program explanations. I offer the following as an olive branch. Given the metaphysically strict analysis of indicative conditionals, $\mathbf{F} \rightarrow \mathbf{X}$ is logically equivalent to $\neg\Diamond(\mathbf{F} \wedge \neg\mathbf{X})$.[3] In all relevant contexts, the truth of this proposition prior to *F*'s instantiation tells us after the instantiation that $\mathbf{X}$ was metaphysically possible - that is, it tells us that it *might have been* that some member of *X* was instantiated. After all, the only time when $\neg\Diamond(\mathbf{F} \wedge \neg\mathbf{X})$ is true but $\mathbf{X}$ is metaphysically impossible is when $\mathbf{F}$ is also metaphysically impossible; that is, when $\mathbf{F} \rightarrow \mathbf{X}$ is vacuously true. I take it that the relevant contexts do not involve explanations in terms of higher-order properties that cannot be instantiated. Hence, in all relevant contexts, the truth of $\mathbf{F} \rightarrow \mathbf{X}$ also tells us how the causal history *might have been*.

What about when the relevant conditional probability is not maximally high? Even though Jackson and Pettit define programming for a causally efficacious property in terms of making the instantiation of some such property 'suitably probable' (1990, p. 114), all of the examples that they give in their paper are ones where the relevant higher-order property '*ensures*... that a crucial productive property is realized'. (1990, p. 114, my italics) In other words, they are all cases where the conditional probability is maximally high. I admit that for any case of program explanation where the relevant conditional probability is not maximally high, the above explanation of the modal information conveyed by the program explanation does not work. But perhaps such cases are simply ones in which no modal information is conveyed. At any rate, it is still good news for the metaphysically strict analysis that the above works as a succinct explanation in cases where the relevant conditional probability is 1, because it still helps us to make sense of Jackson and Pettit's view, which in turn helps us to justify our ordinary practice of giving causal explanations in terms of higher-order, non-efficacious properties.

Before concluding, let us address the question of whether the same succinct explanation can be given on other accounts of conditionals. To make the explanation, one needs two claims. The first claim is that $P(\mathbf{X}|\mathbf{F}) = 1$ *only if* $\mathbf{F} \rightarrow \mathbf{X}$ is true - only this half

[3] In section 6.5, I argued that a proposition of the form $\Diamond(\mathbf{F} \wedge \neg\mathbf{X})$ may be read as a 'might'-conditional. If I am right, then $\neg\Diamond(\mathbf{F} \wedge \neg\mathbf{X})$ may be read 'It is not the case that if *F* is instantiated, then it might be that no member of *X* is instantiated'. This is already a little closer to the language that Jackson and Pettit use.

of McCardel's Thesis is needed to infer $\mathbf{F} \rightarrow \mathbf{X}$. As per the argument for McCardel's Thesis in section 7.7, the claim that $\mathrm{P}(\mathbf{X}|\mathbf{F}) = 1$ only if $\mathbf{F} \rightarrow \mathbf{X}$ is true can be derived on the metaphysically strict analysis given the regularity constraint. According to the regularity constraint, $\mathrm{P}(A) = 0$ only if A is impossible. The sense in which A needs to be impossible depends on the modality expressed by the conditional (assuming that it expresses a modality). For example, given the metaphysically strict analysis, the regularity constraint must convey the *metaphysical* modality; but given a nomically strict analysis, the regularity constraint need only convey the *nomic* modality; and so on. If one adopts the material interpretation, then all one needs in order to justify the claim that $\mathrm{P}(\mathbf{X}|\mathbf{F}) = 1$ only if $\mathbf{F} \rightarrow \mathbf{X}$ is true is a constraint that implies $\mathrm{P}(A) = 0$ only if A is *false*. That is a much easier constraint to justify. But of course, the second claim needed to make the succinct explanation above is that $\mathbf{F} \rightarrow \mathbf{X}$ is a modal proposition. In particular, $\mathbf{F} \rightarrow \mathbf{X}$ needs to tell us, in some relevant sense, that $\mathbf{F} \wedge \neg\mathbf{X}$ is not *possible*. Any analysis that tells us only about the truth value of $\mathbf{F} \wedge \neg\mathbf{X}$ in the actual world or in the nearest possible world(s) will not do - what is needed is a *strict* analysis of some sort. For these reasons, it seems to be a distinctive feature of the view that I endorse that it is able to offer the above explanation of the fact that program explanation conveys the modal information it conveys.

8.5 Conclusion

I began this chapter by explaining Jackson and Pettit's conceptions of causal efficacy and causal explanation. This led to a paradox of causal explanation that was solved by appealing to their view of program explanation. That solution acted as an illustration of a novel perspective according to which program explanation plays a distinct and important role in virtue of conveying modal information. Finally, I explained my novel perspective on that novel perspective. I hope to have provided a simple but effective application of the metaphysically strict analysis by showing that it helps us to make sense of Jackson and Pettit's view, which in turn helps us to justify our ordinary practice of causal explanation.

Chapter 9

Conditional Risk

At first glance, it seems that the risk of an event varies with the probability of its occurrence. For example, the risk of death in a game of Russian roulette seems to vary with the probability of death. However, there has recently been some pushback against probabilistic conceptions of risk in academic philosophy. In particular, Pritchard (2015) argues in favour of a non-probabilistic conception of risk according to which risk varies with comparative similarity between possible worlds. This modal conception of risk is designed to accommodate the idea that the risk of an event can depend on factors other than the probability and disvalue of its occurrence.

In response, Ebert, Smith, and Durbach (2019) argue that Pritchard's modal conception can be supplanted by another non-probabilistic conception of risk on which risk varies not with comparative similarity but with comparative *normalcy* between possible worlds. This normic[1] conception of risk can accommodate the same idea while also avoiding some recalcitrant consequences of Pritchard's modal conception. For this reason, Ebert, Smith, and Durbach argue that *if* one is going to endorse a non-probabilistic conception of risk, then one ought to endorse the normic conception over the modal conception.

In this chapter, I offer a novel argument against Pritchard's modal conception of risk. I do so by drawing on the idea, suggested in section 7.6, that conditional probabilities are more fundamental than unconditional ones. This argument lends some additional credibility to McCardel's Thesis.[2] But more to the point, it demonstrates that the ideas involved in the analysis of conditionals *matter*. I begin by explaining the probabilistic orthodoxy on the nature of risk. I then explain the thought experiment that leads

[1] This word is not to be confused with 'nomic', which (as we saw back in section 3.5) means 'relating to the laws of nature'.

[2] The idea that conditional probability is more fundamental than unconditional probability coheres with The Conditional Principle: an updated version of Lewis's Principal Principle according to which conditional objective chance is rational credence conditional on the right kind of knowledge (see section 7.6). Without this Lewisian view of conditional objective chance, the regularity constraint (see section 7.5) becomes harder to justify, and without the regularity constraint, one cannot justify McCardel's Thesis.

Pritchard to reject this probabilistic orthodoxy, and the modal conception of risk that he offers in its place. I go on to explain three problems with the modal conception, all of which are at least gestured at by Ebert et al., and all of which seem to be avoided by the normic conception of risk. For our purposes, the most important of these is what I call the 'highness problem': on the modal conception, the risk in certain thought experiments regarding intuitively low-risk situations is *maximally high*. Despite ultimately arguing against Pritchard, I suggest a modification of his modal conception which makes explicit the fact that risk is relative to time, and I argue that this time-relative modal conception avoids all three of the problems gestured at by Ebert et al. However, I then argue that the time-relative modal conception has another problem. In light of chapter 7, I suggest re-framing the debate in terms of conditional risk: the risk of one thing *given* another. Once re-framed this way, the following problem becomes apparent: on the time-relative modal conception, the risk in a certain thought experiment regarding an intuitively low-risk situation is maximally high *given* the setup of the thought experiment.[3] I call this the 'conditional highness problem', and I conclude that it gives us reason to reject Pritchard's modal conception of risk.

9.1 The Probabilistic Orthodoxy

One way to get at the nature of risk is to evaluate risk statements. Let us begin by identifying the kind of risk statements that are of relevance to the debate in question. All risk statements are about something undesirable.[4] One way in which risk statements differ is in terms of the types of undesirable things that they are about. Sometimes risk statements are about undesirable *events*. For example:

(1) *There is a risk that this will contaminate your drinking water.*

Here, the undesirable thing is an event in which your drinking water becomes contaminated. Other risk statements are about undesirable *states of affairs*. For example:

(2) *There is a risk that your drinking water is contaminated.*

Here, the undesirable thing is a state of affairs in which your drinking water is contaminated. Of course, this state of affairs can only obtain if there is also an event in which your drinking water becomes contaminated, so the difference between (1) and (2) is subtle. However, it is a difference worth bearing in mind, because Pritchard explicitly sets up the debate in terms of 'risk events'. (2015, p. 437) For this reason, let us focus on those statements of risk that refer to undesirable events as opposed to states of affairs.

[3] By the 'setup' of a thought experiment, I simply mean the information conveyed (explicitly or implicitly) by the vignette depicting the hypothetical situation(s) that are to be entertained by the experimenter.

[4] Or, at least, something undesired. Nothing herein hinges on this distinction, so let us ignore it.

Among the risk statements that are about undesirable events, there are still some that are not of relevance to this debate. For example:

(3) *A risk of hospital admission is that one spreads the flu.*

In referring to this event as a risk, (3) suggests that one may have the propensity to spread the flu. However, (3) gives no *measure* of this propensity - for all (3) says, it may be that hospital admission is extremely risky, or not very risky at all. The same is true of (1). On the other hand, some risk statements do give a measure of propensity. For example:

(4) *The less one washes one's hands, the higher the risk of catching the flu.*

(5) *The risk of death in a game of Russian roulette is one in six.*

It is statements like (4) and (5), as opposed to statements like (3), that are of relevance to this debate.

The most obvious way to measure the propensity of something is in terms of probability, and indeed, in some technical contexts, 'the risk of P' is simply defined to mean 'the probability of P' for some undesirable event that P (see Hansson (2013, §1)). Of course, we also see risk statements in non-technical contexts that seem to have this meaning. For example, it is charitable to interpret (4) as meaning that the *probability* of catching the flu is higher the less one washes one's hands, since this interpretation makes (4) true. Similarly, it is charitable to interpret (5) in terms of the *probability* of death, since the probability of death in a game of Russian roulette with a six-chamber revolver is 1/6.

The above gives us a simple probabilistic conception of risk, but there is also a more complicated probabilistic conception of risk in terms of expectation value. The expectation value of P is just the mathematical product of P's probability and the value of the event that P. For example, if the probability of P is 0.2, and the value of the event that P is 6, then the expectation value of P is 1.2.[5] In some technical contexts, 'the risk of P' is simply defined as meaning 'the expectation value of P' for some undesirable event that P (again, see Hansson (2013, §1)). Some non-technical risk statements may be interpreted this way as well (although there is rarely a reason to prefer this interpretation over the simple probabilistic interpretation in such contexts). Consider (4) again: the (dis)value of catching the flu remains constant regardless of how much one washes one's hands, so the expectation value of catching the flu simply varies in accordance with the probability of catching it. And consider (5) again: if the (dis)value of death is 1 (and it might as well be, since we can choose whatever scale we like), then the

[5] This concept is related to the *expected* value of a course of action, which is obtained by taking an average of the expectation values of all possible outcomes of the course of action.

expectation value of death in a game of Russian roulette with a six-chamber revolver is 1/6.

If risk statements are correctly interpreted either in terms of probability *or* in terms of expectation value, then risk is probabilistic in the following sense: when the (dis)value of the event that P is held constant, the risk of P varies only if the probability of P varies. The orthodox view on risk is that risk is probabilistic in this sense. Call this the 'probabilistic orthodoxy.' In the next section, I explain why Pritchard rejects this probabilistic orthodoxy in favour of a non-probabilistic, modal conception of risk.

9.2 The Modal Conception of Risk

According to Pritchard (2015), the risk of an event can vary even if the probability and disvalue of the event are held constant. Consider the following cases, adapted from Pritchard (2015, p. 441):

case 1 An evil scientist rigs up a bomb in a populated area. The bomb will explode iff a certain set of numbers are drawn in the next national lottery.[6] The probability of this happening is 'fourteen million to one.'

case 2 An evil scientist rigs up a bomb in a populated area. The bomb will explode iff a certain extremely unlikely series of events unfolds.[7] The probability of this happening is 'fourteen million to one.'

The extremely unlikely series of events that Pritchard offers for case 2 is as follows:

> First, the weakest horse in the field at the Grand National, Lucky Loser, must win the race by at least ten furlongs. Second, the worst team remaining in the FA Cup draw, Accrington Stanley, must beat the best team remaining, Manchester United, by at least ten goals. And third, the queen of England must spontaneously choose to speak a complete sentence of Polish during her next public speech. (Pritchard, 2015, p. 441)

The reader may find themselves unable to believe that the probability of the extremely unlikely series of events in case 2 is 1/14,000,000 - as Ebert et al. (2019, pp.

[6] Pritchard does not make both halves of this biconditional explicit. He writes: 'An evil scientist has rigged up a large bomb... The bomb will *only* detonate, however, *if* a certain set of numbers comes up on the next national lottery draw.' (2015, p. 441, my italics) Still, I take the other half of the biconditional to be heavily implied, because the thought experiment makes very little sense unless it is also the case that the bomb will detonate *if* a certain set of numbers comes up on the next national lottery draw.

[7] Again, Pritchard makes only one half of this biconditional explicit, but again, I take the other half to be heavily implied.

5-6) point out, 'The probability... might reasonably be regarded to be far lower'.[8] If so, simply choose some other extremely unlikely (non-random) series of events about which one *can* believe that the probability of it unfolding is 1/14,000,000.

Pritchard's intuition is that the risk of explosion in case 1 is *higher* than the risk of explosion in case 2 even though the disvalue and probability of the explosion are both the same. (Pritchard, 2015, p. 442) This intuition is not universally shared.[9] Indeed, *I* do not share it. Nonetheless, on the basis of this intuition, Pritchard rejects the probabilistic orthodoxy, arguing instead that risk varies in accordance with how *easily* an undesirable event can occur.

We often treat how easily something can occur as synonymous with how *likely* it is to occur, which makes it difficult at first to see how Pritchard's conception of risk is non-probabilistic. However, certain thought experiments can tease the two things apart. For example, the probability of any given lottery ticket winning may be astronomically low, meaning that it is not very likely to win, and yet there is a sense in which each lottery ticket very *easily* could win: all that is required is that 'a few coloured balls... fall in a certain configuration', as Pritchard (2015, p. 442) puts it.

Let us think about this in terms of possible worlds. Suppose all possible worlds are ordered in terms of their comparative similarity to the actual world. Now consider all the losing tickets in the actual world. For each of these tickets, the chance of winning (prior to losing) is very low. Yet, for each ticket, the closest world in which it wins is *extremely close*: everything can be exactly the same except for the configuration in which some balls happen to fall. Thus, Pritchard thinks that the risk of P varies with the closeness of the closest world in which the event that P occurs. More formally:

modal conception The risk at w of the event that P is proportional to the similarity of the most similar P-world to w.

This modal conception of risk allows Pritchard to explain the intuition that the risk in case 1 is higher than the risk in case 2. Let us call the world in which the risk is being measured the 'measurement world', and let us call the worlds in which the risk event occurs the 'occurrence worlds'. Assume that, in both cases, the measurement world is not an occurrence world - that is, it is not a world in which the bomb ends up detonating. In case 1, the most similar world in which the bomb *does* detonate is very similar to the measurement world: some balls just happen to fall in a different configuration. However, in case 2, the closest world in which the bomb explodes is quite different to the measurement world: the weakest horse at the Grand National cannot *just* win by at least ten furlongs, and the worst team in the FA Cup cannot *just*

[8] Indeed, according to Ebert et al., a back-of-the-envelope calculation on the basis of bookmakers' quotes suggests something in the region of '1 in 125,000,000,000'. (2019, p. 6, footnote 8)

[9] See the appendix in Ebert et al. (2019) for data on this.

beat the best team remaining by ten goals, and so on. In other words, the detonation of the bomb in case 2 requires a divergent causal history, whereas the detonation of the bomb in case 1 does not. Hence, even though the undesirable event is equally likely in cases 1 and 2, it can happen more *easily* in case 1 than in case 2.

Despite the fact that I do not share Pritchard's intuition, I think his argument is worth taking seriously. For one thing, there is a surprisingly large minority of people that *do* share Pritchard's intuition.[10] And as Ebert et al. (2019, p. 16) acknowledge, it may be that there are multiple legitimate conceptions of risk.[11] But more importantly, the modal conception makes sense of a common kind of reasoning regarding risk, which Ebert et al. call 'checklist reasoning'. (2019, pp. 6-9) This is the kind of reasoning whereby one concludes that there is a low risk of a disjunctive proposition on the basis that the risk of each of its disjuncts is low. This kind of reasoning is often used in legal contexts (when determining the risk of wrongful conviction), and in cases of *de minimis* risk management more generally (i.e. cases where risk is evaluated against a low but non-zero threshold).[12] The modal conception validates this type of reasoning, since the most similar world in which a disjunction is true is just the most similar world in which one of its disjuncts is true; hence, on the modal conception of risk, the risk of $P \vee Q$ is equal either to the risk of P or to the risk of Q. On the other hand, probabilistic conceptions of risk do not validate this kind of reasoning, since the probability of $P \vee Q$ is equal to the probability of P *plus* the probability of Q.

Of course, there are some contexts in which checklist reasoning is clearly ill-suited. Consider, for example, the following argument:

(I) *There is a low risk that the first chamber contains the bullet.*

(II) *There is a low risk that the second chamber contains the bullet.*

(III) *There is a low risk that the third chamber contains the bullet.*

Therefore,

(IV) *There is a low risk that either the first chamber or the second chamber or the third chamber contains the bullet.*

In a game of Russian roulette where you know that there is one bullet and six chambers, it is clearly a mistake to reason as above. But on the other hand, in a game of Russian roulette where the proportion of bullets to chambers is both random and unknowable, it is less obviously a mistake to reason as above. Suppose you are playing such a game, and you come to be confident (somehow) that (I), (II), and (III) are all true - perhaps you

[10] Again, see the appendix in Ebert et al. (2019) for data on this.

[11] This idea finds support also in Bricker (2018) and Chalmers (2011), the latter of whom argues that conceptual pluralism ought to be the default view in philosophy more generally.

[12] See Ebert et al. (2019, §3) for more information on the use of checklist reasoning in these contexts.

look at the relevant chambers with heat vision goggles and judge (fallibly) that they are empty on the basis of the thermal imaging. In this context, it is less obviously a mistake to infer (IV). Moreover, this second game of Russian roulette is arguably a more appropriate analogy for many real-life situations than the first. As Hansson writes:

> For good or bad, life is usually more like an expedition into an unknown jungle than a visit to the casino. Most of the time we have to deal with dangers without knowing their probabilities, and often we do not even know what dangers we have ahead of us. (Hansson, 2009, p. 427)

I think this suggests that non-probabilistic conceptions of risk are at least worthy of consideration.

Having said all this, the modal conception of risk has some serious problems in its current presentation. Consider again cases 1 and 2, but suppose this time that, in both cases, the world in which the risk is being measured is a world in which the bomb ends up detonating. On this supposition, the similarity of the most similar world to the measurement world, in both cases, is *maximally high*. (Ebert et al., 2019) This is problematic in three different ways: (i) it conflicts with the intuition that the risk of an event does not depend on whether the event ends up taking place; (ii) it conflicts with Pritchard's motivating intuition that the risk in case 1 is higher than the risk in case 2; and (iii) it conflicts with the intuition that the risk in both cases is just not that high. Let us state these more formally as three separate problems:

dependence problem The risk at w of the event that P depends on whether w is a P-world.[13]

sameness problem The risk of the event that P is the same at all P-worlds.

[13] One might disagree that this is an accurate representation of the problem identified by Ebert et al. They write:

> Suppose one is about to drill into a wall in a West Australian house built in the 1970s, and is wondering about the risk that the wall contains asbestos. On the modal conception, if the wall really does contain asbestos, then the risk is maximally high. In this case, there is a maximally similar world–the actual world–in which the wall contains asbestos. If, on the other hand, the wall does not contain asbestos, then, according to the modal conception, the risk will be lower... In any event, on the modal conception it seems that one cannot make a judgment about the *risk* that the wall contains asbestos without taking a view as to whether it *does* contain asbestos. (Ebert et al., 2019, p. 10)

On the face of it, this is not so much about whether the risk of an event depends on its occurrence; rather, it is about the epistemic limitations that the modal conception imposes on someone who is measuring risk. (Thanks to Martin Smith for pushing this point.) However, this objection is framed in terms of the risk of a state of affairs (the wall's containing asbestos), and as above, Pritchard explicitly sets up the debate in terms of *events*, not states of affairs. (2015, p. 437). The modal conception therefore does not suggest any epistemic limitations of the kind to which Ebert et al. make reference. What the modal conception does suggest is that one's epistemic stance on the risk of the event that one drills into asbestos should be informed by one's epistemic stance on whether the wall contains asbestos. But that is surely unobjectionable - the risk that one drills into asbestos is much higher if there is asbestos there to be drilled into.

highness problem The risk of the event that P is maximally high at all P-worlds.

In the next section, I explain the normic conception of risk that Ebert et al. offer as an alternative to the modal conception. As we will see, it seems to avoid all three of these problems.

9.3 The Normic Conception of Risk

In place of the modal conception of risk, Ebert et al. suggest the normic conception of risk. On this conception of risk, rather than ordering possible worlds according to how similar they are to the actual world, one orders them according to how normal they are from the perspective of the actual world. Risk is then determined analogously to before: it varies with closeness to the actual world. Strictly speaking, Ebert et al. do not confine this conception of risk to apply only to the risk of *events*, but since this is what we are interested in here, let us express the normic conception as follows:

normic conception The risk at w of the event that P is proportional to the normalcy of the most normal P-world from the per -spective of w.

This concept of *normalcy* requires some elucidation. Ebert, Smith, and Durbach have a particular conception of normalcy in mind, one that Smith has put to good use in much of his previous work - see e.g. Smith (2010, 2016). On this conception of normalcy, the normal is not necessarily what happens most often. (If it were, the normic conception would just be a probabilistic conception in disguise - one where the relevant probability is determined by frequency.) Rather, Smith argues that the normal enjoys *explanatory privilege*, meaning that it requires less explanation than the abnormal. To illustrate this, he offers the following example:

> [I]t could be true that Tim is normally home by six, even if this occurrence is not particularly frequent. What *is* required is that exceptions to this generalisation are always explicable as exceptions by the citation of independent, interfering factors – his car broke down, he had a late meeting etc. If this condition is met, then the best way to explain Tim's arrival time each day is to assign his arrival by six a *privileged* or *default* status and to *contrastively* explain other arrival times in relation to this default... (Smith, 2010, pp. 15-6)

This much gives us a sense of what might count as a normal event, but one might still wonder what counts as a normal *world*. On this front, Smith offers the following clarification:

> One way to approach the idea of a 'normal world' is as a kind of idealised model *writ large*. On this sort of picture, worlds will count as normal, from w's perspective, to the extent that they approximate simplified, idealised models of w. (Smith, 2016, p. 113)

Having clarified this, let us see whether the normic conception predicts Pritchard's intuition that the risk in case 1 is higher than the risk in case 2. In case 1, the most normal world in which the conditions for the bomb's detonation are met is *very* normal from the perspective of the measurement world - it is a world in which nothing cries out for explanation (except perhaps the evil scientist's actions). In case 2, however, the most normal world that meets the detonation conditions is much less normal from the perspective of the measurement world - the extremely unlikely series of events is, after all, a series of *abnormalities*, each of which cries out for explanation. Hence, the normic conception predicts that the risk in case 1 is higher than the risk in case 2.

The normic conception also seems to avoid the three problems mentioned above. On the normic conception, the occurrence of an event in w has a bearing on the risk of that event at w only if w is the most normal world from its own perspective. But w need not be the most normal world from its own perspective. As Ebert et al. write:

> While the actual world must count as maximally similar to itself, it will not count as maximally normal–after all, the actual world is witness to any number of abnormal events and states of affairs. (Ebert et al., 2019, p. 13)

In other words, even if the measurement world is a world in which the bomb ends up detonating, it might be that the most normal world in which the bomb detonates is some other world besides the measurement world. This undermines any attempt at an analogous argument for the conclusion that the normic conception is afflicted by the dependence problem. And importantly, the sameness problem and the highness problem both hinge on the dependence problem; hence, all three problems seem to be avoided.

In the next section, I develop a time-relative modal conception of risk, and I argue that it also avoids the three problems mentioned above.

9.4 The Time-Relative Modal Conception of Risk

It might be thought that the modal conception of risk can be defended by rejecting what is known as the *strong centring* assumption.[14] When dealing with comparative similarity, this amounts to the assumption that no world is as similar to w as w is to

[14] Thanks to Neil McDonnell for pointing this out.

itself. Perhaps surprisingly, this assumption is not universally accepted. Lewis for one considers rejecting it in favour of *weak centring*: 'The world [w] is one of [the] closest worlds to [w]; but there may be others as well - worlds differing negligibly from [w], so that they come out just as close to [w] as [w] itself.' (1973, p. 29) However, even given weak centring, the degree of similarity of the most similar worlds in cases 1 and 2 will still be maximally high, and to reject both strong *and* weak centring for comparative similarity orderings would be absurd - as Ebert et al. write, '[o]ne thing we can immediately observe is that no world counts as *mor*e similar to the actual world than it is to itself.' (2019, p. 10, my italics)

I think a more promising way to defend the modal conception of risk is to interpret it so that risk is relative to time. This is plausibly what charity demands of us when interpreting Pritchard anyway. (Given the probabilistic orthodoxy, and given the fact that probabilities vary over time, it is generally taken for granted that risk is relative to time.) One might initially be sceptical that the modal conception can accommodate variation in risk over time. After all, if one compares possible worlds in their temporal entirety (i.e. from their beginning to their end), then the closest possible world in which some event occurs will *always* be the closest possible world in which that event occurs - tomorrow's events will not change today's similarity ordering because they have already been taken into account by today's similarity ordering. To relativise risk to time, we need to make it clear that the objects of comparison are not temporally complete possible worlds, but "time-slices" of possible worlds. In particular, I propose that the best way to interpret the modal conception is to take as objects of comparison the *histories* of possible worlds at specified times, where these extend all the way back to the world's beginning (or indefinitely, if the world has no beginning). On this interpretation, to find the risk of an event in a given world, one compares the history of that world to the histories of worlds in which the event occurs. The risk of the event is then proportional to the similarity of the most similar history of a world in which the event occurs.

For this relativisation to work, one has to be careful to distinguish between two different times: the time at which the risk of an event is measured, and the time at which the event itself occurs (or *would* occur). Call the former the 'time of measurement' and the latter the 'time of occurrence.' To find the risk of an event, we compare the history of the measurement world *at the time of measurement* with the histories of occurrence worlds *at the time of occurrence*. Importantly, the time of measurement is usually at least slightly *before* the time of occurrence.

Suppose, for example, that I want to measure the current risk at the actual world of the event that I contract the flu. I compare the current history of the actual world with the histories of worlds in which I contract the flu *up to the time of contraction*. Since these latter histories include events that take place between the time of measurement and the time of occurrence, none of them will be identical to the history of the actual world up to now. Nonetheless, some of them may be very similar to the history of the actual world up to now, in which case the risk of my contracting the flu is very high. More formally:

time-relative modal conception The risk at w of the event that *P* is proportional to the similarity of the most similar history of a *P*-world at the time of occurrence to the history of w at the time of measurement.

Let us see whether this time-relative modal conception of risk predicts Pritchard's intuition that the risk of detonation in case 1 is higher than the risk of detonation in case 2. In case 1, the most similar history of a world in which the bomb detonates is very similar to the history of the measurement world: the histories need only differ in terms of the events that take place between the time of measurement and the time of occurrence, and nothing divergent needs to happen in this period of time. On the other hand, in case 2, the most similar history of a world in which the bomb detonates is less similar to the history of the measurement world: even though the histories may be identical up to the time of measurement, some divergent things need to happen between the time of measurement and the time of occurrence in order for the bomb to go off. In sum, the most similar history of a world in which a few balls fall in an unfortunate configuration is more similar to the history of the measurement world than the most similar history of a world in which the weakest horse at the Grand National wins by at least ten furlongs, etc. Hence, the time-relative modal conception of risk predicts Pritchard's intuition that the risk in case 1 is higher than the risk in case 2.

Moreover, the time-relative modal conception predicts that the risk in case 1 is higher than the risk in case 2 even if the measurement world is an occurrence world. Consider again cases 1 and 2, and suppose this time that the measurement world is a world in which the bomb ends up detonating. Unlike before, the time-relative modal conception of risk does not imply that the risk of detonation is maximally high, because even if the history of the measurement world *at the time of occurrence* is the most similar history of an occurrence world to the history of the measurement world *at the time of measurement*, nonetheless it is not *identical* to it - in both cases, the former history has some events that the latter history lacks. Moreover, in case 1, these events are not divergent, whereas in case 2, they are. Hence, according to the time-relative modal conception, the risk in case 1 is higher than the risk in case 2 even on the supposition that the measurement world is an occurrence world. This means we no longer have any reason to think that the risk at w of the event that *P* depends on whether w is a *P*-world. The time-relative modal conception seems, therefore, to avoid the dependence problem. And as above, the sameness problem and the highness problem both hinge on the dependence problem; hence, all three of these problems seem to be avoided.

Nonetheless, something in the vicinity of the highness problem *seems right*. We might just say that the risk in case 1 is *too* high on the time-relative modal conception of risk.[15] However, I think we can do better than this. In the following section, I re-frame the debate in terms of conditional risk, and I argue that, on the time-relative modal conception, the risk in case 1 is *maximally* high *given* the setup of the thought experiment.

9.5 Against Non-Probabilistic Conceptions of Risk

According to the Lewisian view of objective chance espoused in chapter 7, objective chance is rational credence conditional on the right kind of knowledge - or rather, *conditional* objective chance is rational credence conditional on the right kind of knowledge. This is the idea captured by what I called 'The Conditional Principle' - an updated version of Lewis's Principal Principle, explained in section 7.6. Therein, I suggested that conditional probabilities might be thought to be more fundamental than unconditional ones. With this in mind, we should be asking the following question: 'What is the *conditional* probability of the bomb's detonation in cases 1 and 2 *given* the setup of the thought experiment?' On a simple probabilistic conception of risk, the answer to this question tells us the conditional risk of the bomb's detonation in cases 1 and 2 *given* the setup of the thought experiment. I suggest it is this - not *un*conditional risk - that we should be interested in first and foremost when comparing different conceptions of risk. Of course, our judgements of risk in cases 1 and 2 are already based on the setup of the thought experiment, so it may surprise the reader to find that re-framing the debate this way makes a difference. Yet it does make a difference. In particular, it draws our attention to a problem regarding the time-relative modal conception that is otherwise easily missed:

conditional highness problem The risk in case 1 is maximally high *given* the setup of the thought experiment.

Let us assume that it is implicit in the setup of case 1 that the time of measurement is prior to the time of occurrence. Thus, the setup of case 1 ensures that no history of an occurrence world at the time of occurrence is maximally similar to the history of the measurement world at the time of measurement. Nonetheless, the setup of case 1 is still consistent with a *range* of degrees of similarity: some histories of occurrence worlds

[15] Thanks to Adam Rieger for pointing this out to me.

are more similar to the history of the measurement world than others. My contention is that the similarity of the most similar history of an occurrence world in case 1 is of the highest degree *within that range*. After all, the most similar history of a world in which the risk event occurs is just as similar to the history of the measurement world as the most similar history in which the risk event does *not* occur: they both involve some balls falling in apparently random configurations, and they both end at the time of occurrence. In sum: even if the degree of similarity of the most similar history of an occurrence world is not maximally high *full stop*, it *is* maximally high *given* the setup.

Let us formalise this idea to make it clearer. Take all the sets of balls that might be drawn in the lottery and label them 'set 1', 'set 2', 'set 3', etc. Let X be the set of events comprised of the event that set 1 is drawn, the event that set 2 is drawn, the event that set 3 is drawn, etc. Lastly, let $\mathbf{X}$ be the proposition that some member of X occurs, and let $\mathbf{S}$ be the proposition that the relevant setup is true. To say that something is maximally high is to say that it is as high as it could be, and as we saw back in chapter 3, words like 'could' are ambiguous between different types of modality. My contention is that the relevant modality in this context encompasses all and only those metaphysically possible worlds at which the setup is true, i.e. $\mathbf{S}$-worlds. Now suppose set 13 is the unfortunate set, meaning the risk event in case 1 occurs iff set 13 is drawn. The trouble is that the most similar history of an $\mathbf{S}$-world in which set 13 is drawn is no less similar than the most similar history of an $\mathbf{S}$-world in which some other member of X occurs. Hence, even if the degree of similarity of the most similar history in which the risk event occurs is not maximally high *full stop*, it *is* maximally high *given* $\mathbf{S}$.

No such problem arises in case 2, because the most similar history of a world in which the risk event occurs is much less similar than the most similar history of a world in which the risk event does not occur. Nonetheless, we can expect the same problem to arise in any thought experiment in which a risk event occurs iff one of a set of equally similar histories unfolds. I think this gives us reason to reject the time-relative modal conception of risk. Put simply, it is a wild overstatement to say that the risk in case 1 is as high as it could be given the setup. (If this statement sounds confused, I sympathise. More on this shortly.)

Importantly, the conditional highness problem does not afflict either of the probabilistic conceptions of risk discussed above. As above, Pritchard claims that the probability of the risk event's occurrence in both cases is 'fourteen million to one.' (2015, p. 441) This figure was presumably obtained by rounding to the nearest million given a lottery in which six balls are drawn from 49. (The probability of any particular six-ball set in a fair lottery with 49 balls is 1/13,983,816.)[16] However, Pritchard also claims that 'in both cases the chances of [the risk event] occurring are by stipulation identical.'

[16] $49! \div 6!(49-6)! = 13,983,816.$

(2015, p. 441) That requires that the probability in both cases is not just approximate, but *exact*. The setup thus suffices to determine the probability of the risk event in case 1. In sum: while the setup may be consistent with a range of degrees of similarity, it is not consistent with a range of "degrees of likelihood." And because of this, there does not seem to be any sense in which the probability of the risk event is *maximally* high given the setup. (This explains why the statement 'the risk in case 1 is as high as it could be given the setup' may sound confused - on a simple probabilistic conception of risk, it *is* confused.)

What if Pritchard had simply said that the probability in each case is identical without specifying exactly what that probability is, thereby making the setup of case 1 consistent with a range of probabilities? In that case, the probabilistic conceptions of risk discussed in section 9.1 simply do not make any predictions on the basis of the setup of case 1 (assuming that it is not implicit in the setup that the lottery involves six balls being drawn from 49). Therefore, even given an alternative vignette, the probabilistic conceptions do not predict that the risk in case 1 is maximally high within the range of probabilities consistent with the setup.

Interestingly, it is not clear whether the conditional highness problem afflicts the normic conception of risk. Just as the setup of case 1 is consistent with a range of degrees of similarity, so it is consistent with a range of degrees of *normalcy*: some occurrence worlds are more normal from the perspective of the measurement world than others. However, it is not clear that the normalcy of the most normal occurrence world is of the highest degree within that range. The question is: Is the most normal **S**-world in which the risk event occurs just as normal from the perspective of the measurement world as the most normal **S**-world in which the risk event does not occur? I do not think we are in a position to answer that question - we do not know enough about the measurement world to gauge what is normal from its perspective. Perhaps we can say that the most normal *history* of an occurrence world (up to the time of occurrence) is just as normal as the most normal *history* of a world in which the detonation does not occur (up to the same time). But on the normic conception of risk, we have to take into consideration the events that occur after the time of occurrence as well, and it just is not clear how normal these events are from the perspective of the measurement world in case 1. This makes the normic conception a bit impotent, but it does not make it *wrong*. (Besides, as per Hansson's quote above, there are many cases in which probabilistic conceptions are a bit impotent as well.) For all I have said, then, the normic conception may be a legitimate conception of risk. If so, I suggest we follow Bricker (2018) and Ebert et al. (2019) in advocating a *plurality* of legitimate conceptions of risk, excluding the time-relative modal conception.

9.6 Conclusion

In this chapter, I have used the idea that conditional probabilities are more fundamental than unconditional ones to argue against Pritchard's non-probabilistic conception of risk. I began by explaining the probabilistic orthodoxy on the nature of risk. I then explained two non-probabilistic conceptions of risk: the modal conception and the normic conception. In section 9.4, I argued that the modal conception ought to be modified so as to be compatible with the idea that risk is relative to time. (Plausibly, this is what Pritchard had in mind anyway.) Finally, I re-framed the debate in terms of conditional risk and argued that Pritchard's non-probabilistic conception suffers from what I have called the 'conditional highness problem': the problem of predicting a maximally high risk given the setup of a thought experiment about an intuitively low-risk situation. The ideas defended in chapter 7 are thus instrumental in clarifying the nature of risk.

Chapter 10

Conclusion

Over the course of this thesis, I have tried to show that the metaphysically strict analysis of indicative conditionals is a serious contender among rivals. I hope to have achieved this, not by making short and snappy arguments in its favour, but by fleshing out a coherent wider view that draws on some of the philosophical developments of the last century. I began by arguing that it would be good to have an analysis of indicative conditionals according to which they express propositions, and that a modal analysis meets this requirement while avoiding some of the pitfalls of the material interpretation. More specifically, I argued that both the metaphysically strict analysis and the nomically strict analysis avoid making any intuitively true propositions logically impossible. I then argued that the former has a slight edge over the latter given a Humean best-system analysis of the laws of nature. In chapter 4, I defended the metaphysically strict analysis against the objection that it makes too many intuitively true conditionals false by arguing that many false conditionals are nonetheless assertable, and that we sometimes confuse assertability with truth. Then, in chapter 5, I argued that dualism regarding conditionals should not be concluded on the basis of arguments from correspondence, and that the intuitive difference between corresponding indicative and counterfactual conditionals can be explained in terms of assumptions relating to assertability conditions. In chapter 6, I argued that the metaphysically strict analysis can be extended to 'would'-conditionals to form a unified theory of natural language conditionals in general - except 'might'-conditionals, which are statements of possibility rather than statements of necessity. In chapter 7, I argued that the metaphysically strict analysis allows us to draw a connection between conditionals and conditional probability by invoking a Lewisian view of conditional objective chance. In chapter 8, I argued that this helps us to succinctly explain Jackson and Pettit's view of program explanation when combined with the metaphysically strict analysis. And in chapter 9, I used the idea that conditional probability is more fundamental than unconditional probability to offer a novel argument against a non-probabilistic conception of risk.

The above gives a detailed and precise summary of the previous nine chapters, but a less detailed and precise summary may help to get the main points across. Here is one such summary. The main cost of the material interpretation is that too many conditionals are true. One can try to fix this by making conditionals subjective or non-truth-evaluable, or one can just accept that the truth conditions of conditionals diverge from intuition. I think the latter option is preferable. Having accepted this, one can either endorse a simple, material interpretation of indicative conditionals *and* a more complicated, modal analysis of counterfactual conditionals, *or* one can endorse a unified, modal analysis of both types of conditional, explaining their difference by appeal to context. I think the latter option is preferable. One can then either place this context-sensitivity in the truth conditions of conditionals as per Stalnaker's theory, or one can place it in the assertability conditions where context is already commonly believed to have an impact. I think the latter option is preferable.

The resultant view of conditionals is one that has many advantages over its competitors: it avoids paradoxes of material implication; it pays at least *minimal* respect to our intuitions; it is compatible with the idea that conditionals express propositions; it allows us to draw a plausible connection between conditionals and conditional probability; it helps us to understand causal explanation; it coheres with the fruitful idea that conditional probability is more fundamental than unconditional probability; it even helps us to make sense of the fact that 'if and only if' is the standard way of expressing necessary and sufficient conditions. The main problem with the view is just that its truth conditions diverge from intuition, but even this can be mitigated by plugging in different metaphysical theories: perhaps necessitarianism regarding the laws of nature is true, or perhaps strict necessitarianism is true. Overall, this should suffice to encourage contemporary philosophers to treat the metaphysically strict analysis as a serious contender among rivals.

Before finishing, there are still some questions that could do with being answered. For example: 'Do metaphysically strict conditionals obey the principle of Monotonicity?' This is the principle, also known as 'strengthening the antecedent', according to which any true conditional implies another true conditional with a logically stronger antecedent:

Monotonicity $(A > C) \supset ((A \wedge B) > C)$.

Answer: 'Yes'. After all, if it is metaphysically impossible that A and not-C, then it is also metaphysically impossible that A and B and not-C.

Another question: 'Do metaphysically strict conditionals obey the principle of Transitivity?' This is the principle according to which the following chain of reasoning is valid:

Transitivity $(A > B) \supset ((B > C) \supset (A > C))$.

Answer: 'Yes'. This can be proven using even the weakest modal logic system, known as 'K'. To prove it, we interpret '>' as a metaphysically strict conditional connective, giving us:

transitivity hypothesis $\Box(A \supset B) \supset (\Box(B \supset C) \supset \Box(A \supset C))$.

We then suppose that this hypothesis is false and see whether or not we are led to absurdity. The tableau[1] below shows that we *are* led to absurdity; hence, the transitivity hypothesis is true.

1.	$\neg(\Box(A \supset B) \supset (\Box(B \supset C) \supset \Box(A \supset C))), \mathrm{w}_1$ ✓	hypothesis negation
2.	$\Box(A \supset B), \mathrm{w}_1$ ✓	$1 \neg \supset$ rule
3.	$\neg(\Box(B \supset C) \supset \Box(A \supset C)), \mathrm{w}_1$ ✓	$1 \neg \supset$ rule
4.	$\Box(B \supset C), \mathrm{w}_1$ ✓	$3 \neg \supset$ rule
5.	$\neg\Box(A \supset C), \mathrm{w}_1$ ✓	$3 \neg \supset$ rule
6.	$\Diamond\neg(A \supset C), \mathrm{w}_1$ ✓	$5 \neg\Box$ rule
7.	$\mathrm{w}_1 R \mathrm{w}_2$ ✓	$6 \Diamond$ rule
8.	$\neg(A \supset C), \mathrm{w}_2$ ✓	$6 \Diamond$ rule
9.	A, w_2	$8 \neg \supset$ rule
10.	$\neg C, \mathrm{w}_2$	$8 \neg \supset$ rule
11.	$A \supset B, \mathrm{w}_2$ ✓	$2,7 \Diamond$ rule
12.	$\neg A, \mathrm{w}_2$ B, w_2	$11 \supset$ rule
13.	$\otimes$ (9, 12) $B \supset C, \mathrm{w}_2$ ✓	$4,7 \Box$ rule
14.	$\neg B, \mathrm{w}_2$ C, w_2	$13 \supset$ rule
	$\otimes$ (12, 14) $\otimes$ (10, 14)	

In a nutshell, what this tableau tells us is that there is no way of making the formula in

[1] Line 1 of the tableau represents the supposition that the transitivity hypothesis is false. Lines 2 and 3 are derived from line 1 together with the fact that a material conditional is false iff its antecedent is true and its consequent is false. Lines 4 and 5 are derived from the same fact together with line 3. Line 6 is derived from the fact that what is not necessary is possibly false (see definitions in section 2.3). Line 7 is a representation of the fact that there must be a world, w_2, which is accessible from w_1, since *something* is possible in w_1. (No assumptions are made about w_2 - it may be the same world as w_1.) Line 8 represents the *something* that is possible in w_2. Lines 9 and 10 are derived from line 8 together with the fact that a material conditional is false iff its antecedent is true and its consequent is false. Line 11 is derived from lines 2 and 7 together with the fact that what is necessary is true in all possible worlds, including w_2. Line 12 is derived from the fact that a material conditional is true iff its antecedent is false or its consequent is true. The left hand branch of the tree then closes due to inconsistency: it features both A and $\neg A$ in w_2, on lines 9 and 12 respectively. The right hand branch continues to line 13, which is derived from lines 4 and 7 together with the fact that what is necessary is true in all possible worlds. Lastly, line 14 is derived from the fact that a material conditional is true iff its antecedent is false or its consequent is true. Both remaining branches then close due to inconsistency, as indicated.

line 1 true. And that means that the formula's negation (i.e. the transitivity hypothesis) is a necessary truth. Hence, conditionals are transitive on the metaphysically strict analysis.

Another question: 'Do metaphysically strict conditionals obey the principle of Contraposition?' This is the principle, first encountered in section 2.5, according to which swapping and negating a conditional's antecedent and consequent preserves truth value:

Contraposition $P > Q \equiv \neg Q > \neg P.$

Answer: 'Yes'. It is metaphysically impossible that P and not-Q iff it is metaphysically impossible that not-Q and P.

There are of course some counterexamples to contraposition. To borrow one from Edgington (2009, §2), if we contrapose the conditional 'If it rains, it won't rain much', we get 'If it rains much, it won't rain.' The first conditional seems assertable, whereas the second seems absurd. However, this is perfectly consistent with the view of conditionals that I have defended. As per chapter 4, the assertability of the first conditional does not imply that it is true; hence, we can maintain that both conditionals are false. And we have no reason to think that contraposition preserves assertability.

Questions regarding the logical treatment of nested conditionals are somewhat trickier to answer. For example: 'Do metaphysically strict conditionals obey the principle of Import-Export?' This is the principle according to which a right-nested conditional is equivalent to a non-nested conditional:

Import-Export $A > (B > C) \equiv (A \wedge B) > C.$

The answer seems to be: 'No'. Here is an apparent counterexample: A, B, and C are all metaphysically contingent, but B is equivalent to $\neg A$. In that case, the non-nested conditional is vacuously true (because it has a metaphysically impossible antecedent), whereas the nested conditional may easily be false.

However, I am reluctant to draw any firm conclusions about how to interpret nested conditionals. It is not at all clear that simple nested modal propositions serve a purpose in natural language, and as chapter 6 demonstrates, we should not assume that the logical structure of a proposition can simply be read off from the structure of the sentence expressing it. It may be that 'If it is going to rain tomorrow afternoon, then if we want to go to the park, we should go in the morning' really just expresses the same proposition as 'If it is going to rain tomorrow afternoon and we want to go to the park, then we should go in the morning.'

No doubt other loose ends remain, but that is to be expected when theorising about a topic with far-reaching consequences. And the consequences of this topic *are* far-reaching, even if subtle: conditionals play a prominent role in many areas of inquiry, and the analysis of conditionals helps us to fine-tune our use of them in those areas. More than this, I think the analysis of conditionals helps us to fine-tune that 'basic part of our mental equipment' (to repeat a line from Edgington (1995, p. 235)): 'the ability to think conditional thoughts'. Of course, competent language users already have an intuitive understanding of conditionals, and in many contexts, an intuitive understanding is good enough. But as we have seen, intuition cannot always be trusted, and in some of the most interesting contexts, thinking the right thought matters.

Bibliography

Adams, E. W. (1965). The logic of conditionals. *Inquiry*, *8*(1-4), 166-197. doi: 10.1080/00201746508601430

Adams, E. W. (1970). Subjunctive and indicative conditionals. *Foundations of Language*, *6*(1), 89–94. doi: 10.2307/2272204

Adams, E. W. (1975). *The logic of conditionals: An application of probability to deductive logic*. D. Reidel Publishing Company.

Appiah, A. (1985). *Assertion and conditionals*. Cambridge University Press.

Barrett, J. (2023). Everettian Quantum Mechanics. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy* (Summer 2023 ed.). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2023/entries/qm-everett/.

Bennett, J. (1988). Farewell to the phlogiston theory of conditionals. *Mind*, *97*(388), 509–527. doi: 10.1093/mind/xcvii.388.509

Bennett, J. (1996). Spinoza's metaphysics. In D. Garrett (Ed.), *The Cambridge companion to Spinoza* (pp. 61–88). Cambridge University Press.

Bennett, J. (2003). *A philosophical guide to conditionals*. Oxford University Press.

Bigelow, J., & Pargetter, R. (1990). *Science and necessity* (R. Pargetter, Ed.). Cambridge University Press.

Bird, A. (2005). The dispositionalist conception of laws. *Foundations of Science*, *10*(4), 353–370. doi: 10.1007/s10699-004-5259-9

Blackburn, S. (1986). How can we tell whether a commitment has a truth condition? In C. Travis (Ed.), *Meaning and interpretation* (pp. 201–232). Blackwell Publishing.

Bricker, A. M. (2018). Do judgements about risk track modal ordering? *Thought: A Journal of Philosophy*, *7*(3), 200–208. doi: 10.1002/tht3.388

Carnap, R. (1950). *Logical foundations of probability*. University of Chicago Press.

Chalmers, D. J. (2011). Verbal disputes. *Philosophical Review*, *120*(4), 515–566. doi: 10.1215/00318108-1334478

Chalmers, D. J., & Hájek, A. (2007). Ramsey + Moore = God. *Analysis*, *67*(2), 170–172. doi: 10.1093/analys/67.2.170

Chisholm, R. M. (1946). The contrary-to-fact conditional. *Mind*, *55*(220), 289–307. doi:

10.1093/mind/LV.219.289

Davis, W. A. (1979). Indicative and subjunctive conditionals. *Philosophical Review*, *88*(4), 544–564. doi: 10.2307/2184844

Davis, W. A. (1980). Lowe on indicative and counterfactual conditionals. *Analysis*, *40*(4), 184–186. doi: 10.1093/analys/40.4.184

Douven, I. (2006). Assertion, knowledge, and rational credibility. *Philosophical Review*, *115*(4), 449–485. doi: 10.1215/00318108-2006-010

Dretske, F. I. (1970). Epistemic operators. *Journal of Philosophy*, *67*(24), 1007–1023. doi: 10.2307/2024710

Dudman, V. H. (2000). Classifying 'conditionals': The traditional way is wrong. *Analysis*, *60*(2), 147–147. doi: 10.1111/1467-8284.00216

Dummett, M. (1973). *Frege: Philosophy of language*. Duckworth.

Ebert, P. A., Smith, M., & Durbach, I. (2019). Varieties of risk. *Philosophy and Phenomenological Research*, 1–24. doi: 10.1111/phpr.12598

Edgington, D. (1986). Do conditionals have truth conditions? *Critica*, *18*(52), 3–39.

Edgington, D. (1995). On conditionals. *Mind*, *104*(414), 235–329. doi: 10.1093/mind/104.414.235

Edgington, D. (2009). Conditionals, truth and assertion. In I. Ravenscroft (Ed.), *Minds, ethics, and conditionals: Themes from the philosophy of Frank Jackson.* Oxford University Press.

Edgington, D. (2020). Indicative Conditionals. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Fall 2020 ed.). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2020/entries/conditionals/.

Ellis, B. (1969). An epistemological concept of truth. In R. Brown & C. D. Rollins (Eds.), *Contemporary philosophy in Australia* (pp. 52–72). George Allen & Unwin Ltd.

Ellis, B. (1978). A unified theory of conditionals. *Journal of Philosophical Logic*, *7*(1), 107–124. doi: 10.1007/bf00245924

Ellis, B. (2001). *Scientific essentialism.* Cambridge University Press.

Fales, E. (1993). Are causal laws contingent? In J. Bacon, K. Campbell, & L. Reinhardt (Eds.), *Ontology, causality and mind: Essays in honour of D.M. Armstrong.* Cambridge University Press.

Fine, K. (2002). Varieties of necessity. In T. S. Gendler & J. Hawthorne (Eds.), *Conceivability and possibility* (pp. 253–281). Oxford University Press.

Fine, K. (2006). Relatively unrestricted quantification. In A. Rayo & G. Uzquiano (Eds.), *Absolute generality* (pp. 20–44). Oxford University Press.

Fraassen, B. V. (1966). Singular terms, truth-value gaps, and free logic. *Journal of Philosophy*, *63*(17), 481–495. doi: 10.2307/2024549

Garrett, D. (1991). Spinoza's necessitarianism. In Y. Yovel (Ed.), *God and nature: Spinoza's* *metaphysics* (pp. 191–218). Brill.

Gazdar, G. (1979). *Pragmatics: Implicature, presupposition and logical form*. Academic Press.

Gibbard, A. (1980a). Indicative conditionals and conditional probability: Reply to Pollock. In W. L. Harper, R. C. Stalnaker, & G. Pearce (Eds.), *Ifs: Conditionals, belief, decision, chance, and time* (pp. 253–256). D. Reidel Publishing Company.

Gibbard, A. (1980b). Two recent theories of conditionals. In W. L. Harper, R. C. Stalnaker, & G. Pearce (Eds.), *Ifs: Conditionals, belief, decision, chance, and time* (pp. 211–247). D. Reidel Publishing Company.

Goldfarb, W. (2003). *Deductive logic*. Hackett Publishing Company.

Goldman, A. I. (1976). Discrimination and perceptual knowledge. *Journal of Philosophy*, *73*(20), 771–791. doi: 10.2307/2025679

Goodman, N. (1947). The problem of counterfactual conditionals. *Journal of Philosophy*, *44*(5), 113–128. doi: 10.2307/2019988

Gozzano, S. (2020). Necessitarianism and dispositions. *Metaphysica*, *21*(1), 1–23. doi: 10.1515/mp-2019-0022

Grice, H. P. (1967a). Indicative conditionals. In H. P. Grice (Ed.), *Studies in the way of words* (pp. 58–85). Harvard University Press.

Grice, H. P. (1967b). Logic and conversation. In H. P. Grice (Ed.), *Studies in the way of words* (pp. 22–40). Harvard University Press.

Griffin, M. V. (2012). Necessitarianism in Spinoza and Leibniz. In *Leibniz, God and necessity* (p. 58–82). Cambridge University Press.

Hájek, A. (2003). What conditional probability could not be. *Synthese*, *137*(3), 273–323. doi: 10.1023/b:synt.0000004904.91112.16

Hájek, A. (2011). *Staying regular*. [Unpublished manuscript].

Hájek, A. (2012a). The fall of "Adams' thesis"? *Journal of Logic, Language and Information*, *21*(2), 145–161. doi: 10.1007/s10849-012-9157-1

Hájek, A. (2012b). Is strict coherence coherent? *Dialectica*, *66*(3), 411–424. doi: 10.1111/j.1746-8361.2012.01310.x

Hájek, A. (2014). *Most counterfactuals are false*. [Unpublished manuscript].

Hansson, S. O. (2009). From the casino to the jungle: Dealing with uncertainty in technological risk management. *Synthese*, *168*(3), 423–432. doi: 10.1007/s11229-008-9444-1

Hansson, S. O. (2013). *The ethics of risk: Ethical analysis in an uncertain world*. Palgrave Macmillan.

Huddleston, R., Pullum, G. K., & Reynolds, B. (2021). *A student's introduction to English grammar* (2nd ed.). Cambridge University Press.

Jackson, F. (1979). On assertion and indicative conditionals. *Philosophical Review*, *88*(4), 565–589.

Jackson, F. (1990). Classifying conditionals. *Analysis*, *50*(2), 134–147. doi: 10.1093/analys/50.2.134

Jackson, F. (1991). *Conditionals*. Oxford University Press.

Jackson, F., & Pettit, P. (1988). Functionalism and broad content. *Mind*, *97*(387), 318–400. doi: 10.1093/mind/XCVII.387.381

Jackson, F., & Pettit, P. (1990). Program explanation: A general perspective. *Analysis*, *50*(2), 107–17. doi: 10.1093/analys/50.2.107

Jeffrey, R. C. (1964). If (abstract). *Journal of Philosophy*, *61*, 702–703.

Jeffrey, R. C. (1965). *The logic of decision*. University of Chicago Press.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kasper, W. (1992). Presuppositions, composition, and simple subjunctives. *Journal of Semantics*, *9*(4), 307–331. doi: 10.1093/jos/9.4.307

Kelp, C. (2018). Assertion: A function first account. *Noûs*, *52*(2), 411–442. doi: 10.1111/nous.12153

Kment, B. (2021). Varieties of Modality. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2021 ed.). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2021/entries/modality-varieties/.

Kolmogorov, A. (1956). *Foundations of the theory of probability*. Chelsea Publishing Company.

Kripke, S. (1980). *Naming and necessity*. Harvard University Press.

Lackey, J. (2007). Norms of assertion. *Noûs*, *41*(4), 594–626. doi: 10.1111/j.1468-0068.2007.00664.x

Lepage, F. (2015). Is Lewis's triviality result actually a triviality result? *Logique et Analyse*, *58*(231), 371–375.

Lewis, C. I. (1912). Implication and the algebra of logic. *Mind*, *21*(84), 522–531. doi: 10.1093/mind/xxi.84.522

Lewis, C. I. (1914). The calculus of strict implication. *Mind*, *23*(90), 240–247.

Lewis, C. I. (1918). *A survey of symbolic logic*. University of California Press.

Lewis, D. K. (1968). Counterpart theory and quantified modal logic. *Journal of Philosophy*, *65*(5), 113–126. doi: 10.2307/2024555

Lewis, D. K. (1973). *Counterfactuals*. Blackwell Publishing.

Lewis, D. K. (1976). Probabilities of conditionals and conditional probabilities. *Philosophical Review*, *85*(3), 297–315. doi: 10.2307/2184279

Lewis, D. K. (1980). A subjectivist's guide to objective chance. In R. C. Jeffrey (Ed.), *Studies in inductive logic and probability, volume ii* (pp. 263–293). University of California Press.

Lewis, D. K. (1986a). *On the plurality of worlds*. Wiley-Blackwell.

Lewis, D. K. (1986b). Probabilities of conditionals and conditional probabilities ii. *Philosophical Review*, *95*(4), 581–589. doi: 10.2307/2185051

Lewis, D. K. (1994). Humean supervenience debugged. *Mind*, *103*(412), 473–490. doi: 10.1093/mind/103.412.473

Lowe, E. J. (1979). Indicative and counterfactual conditionals. *Analysis*, *39*(3), 139–141. doi: 10.1093/analys/39.3.139

Lowe, E. J. (1980). Reply to Davis. *Analysis*, *40*(4), 187–190. doi: 10.1093/analys/40.4.187

Lycan, W. G. (2001). *Real conditionals*. Oxford University Press.

MacColl, H. (1880). Symbolical reasoning. *Mind*, *5*(17), 45–60.

McKinnon, R. (2015). *Norms of assertion: Truth, lies, and warrant*. Palgrave-Macmillan.

Menzel, C. (2023a). The Possibilism-Actualism Debate. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy* (Spring 2023 ed.). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2023/entries/possibilism-actualism/.

Menzel, C. (2023b). Possible Worlds. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy* (Summer 2023 ed.). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2023/entries/possible-worlds/.

Nolan, D. (1997). Impossible worlds: A modest approach. *Notre Dame Journal of Formal Logic*, *38*(4), 535–572. doi: 10.1305/ndjfl/1039540769

Peirce, C. S. (1896). The regenerated logic. *The Monist*, *7*(1), 19–40. doi: 10.5840/monist18967121

Popper, K. (1959). *The logic of scientific discovery*. Basic Books.

Pritchard, D. (2015). Risk. *Metaphilosophy*, *46*(3), 436–461. doi: 10.1111/meta.12142

Quine, W. V. O. (1950). *Methods of logic*. Harvard University Press.

Quine, W. V. O. (1960). *Word and object*. MIT Press.

Ramsey, F. P. (1931a). General propositions and causality. In F. P. Ramsey (Ed.), *The foundations of mathematics and other logical essays* (pp. 237–255). Routledge and Kegan Paul.

Ramsey, F. P. (1931b). Truth and probability. In F. P. Ramsey (Ed.), *The foundations of mathematics and other logical essays* (pp. 156–198). Routledge and Kegan Paul.

Read, S., & Edgington, D. (1995). Conditionals and the Ramsey test. *Aristotelian Society Supplementary Volume*, *69*(1), 47–86. doi: 10.1093/aristoteliansupp/69.1.47

Rieger, A. (2006). A simple theory of conditionals. *Analysis*, *66*(3), 233–240. doi: 10.1093/analys/66.3.233

Rieger, A. (2013). Conditionals are material: The positive arguments. *Synthese*, *190*(15), 3161–3174. doi: 10.1007/s11229-012-0134-7

Ryle, G. (1949). *The concept of mind*. Hutchinson & Co.

Shimony, A. (1955). Coherence and the axioms of confirmation. *Journal of Symbolic*

Logic, *20*(1), 1–28. doi: 10.2307/2268039

Shoemaker, S. (1980). Causality and properties. In P. van Inwagen (Ed.), *Time and cause* (pp. 109–35). D. Reidel Publishing Company.

Shoemaker, S. (1998). Causal and metaphysical necessity. *Pacific Philosophical Quarterly*, *79*(1), 59–77. doi: 10.1111/1468-0114.00050

Sider, T. (2010). *Logic for philosophy*. Oxford University Press.

Simion, M. (2016). Assertion: Knowledge is enough. *Synthese*, *193*(10). doi: 10.1007/s11229-015-0914-y

Skyrms, B. (1995). Strict coherence, sigma coherence and the metaphysics of quantity. *Philosophical Studies*, *77*(1), 39–55. doi: 10.1007/bf00996310

Smith, M. (2010). What else justification could be. *Noûs*, *44*(1), 10–31. doi: 10.1111/j.1468-0068.2009.00729.x

Smith, M. (2016). *Between probability and certainty: What justifies belief*. Oxford University Press.

Stalnaker, R. C. (1968). A theory of conditionals. In N. Rescher (Ed.), *Studies in logical theory* (pp. 98–112). Blackwell Publishing.

Stalnaker, R. C. (1970a). Pragmatics. *Synthese*, *22*(1-2), 272–289. doi: 10.1007/bf00413603

Stalnaker, R. C. (1970b). Probability and conditionals. *Philosophy of Science*, *37*(1), 64–80. doi: 10.1086/288280

Stalnaker, R. C. (1981). A defense of conditional excluded middle. In W. L. Harper, R. C. Stalnaker, & G. Pearce (Eds.), *Ifs: Conditionals, belief, decision, chance, and time* (pp. 87–104). D. Reidel Publishing Company.

Stalnaker, R. C. (1999). Assertion. In *Context and content: Essays on intentionality in speech and thought* (pp. 78–95). Oxford University Press.

Starr, W. B. (2014). A uniform theory of conditionals. *Journal of Philosophical Logic*, *43*(6), 1019–1064. doi: 10.1007/s10992-013-9300-8

Swoyer, C. (1982). The nature of natural laws. *Australasian Journal of Philosophy*, *60*(3), 203–223. doi: 10.1080/00048408212340641

Weiner, M. (2005). Must we know what we say? *Philosophical Review*, *114*(2), 227–251. doi: 10.1215/00318108-114-2-227

Whitehead, A. N., & Russell, B. (1910). *Principia mathematica*. Cambridge University Press.

Williamson, T. (1996). Knowing and asserting. *Philosophical Review*, *105*(4), 489–523. doi: 10.2307/2998423

Williamson, T. (2007). How probable is an infinite sequence of heads? *Analysis*, *67*(3), 173–180. doi: 10.1111/j.1467-8284.2007.00671.x

Williamson, T. (2020). *Suppose and tell: The semantics and heuristics of conditionals*. Oxford University Press.

Woods, M. (1997). *Conditionals* (D. Wiggins & D. Edgington, Eds.). Oxford University Press.

Woodward, J. F. (2003). *Making things happen: A theory of causal explanation*. Oxford University Press.

Yagisawa, T. (1988). Beyond possible worlds. *Philosophical Studies*, *53*(2), 175–204. doi: 10.1007/bf00354640