THE UNIVERSITY OF QUEENSLAND

AUSTRALIA

Context-indexed Counterfactuals and Non-vacuous Counterpossibles.

Mariusz Popieluch

BA with majors in Philosophy and Mathematics

With Honours (Class I) in Philosophy

A thesis submitted for the degree of Doctor of Philosophy at

The University of Queensland in 2018

School of Historical and Philosophical Inquiry

Abstract

The two main features of this thesis are (i) an account of contextualized (context indexed) counterfactuals, and (ii) a non-vacuist account of counterpossibles. Experience tells us that the truth of the counterfactual is contingent on what is meant by the antecedent, which in turn rests on what context is assumed to underlie its reading (intended meaning). On most conditional analyses, only the *world of evaluation* and the *antecedent* determine which worlds are relevant to determining the truth of a conditional, and consequently what its truth value is. But that results in the underlying context being fixed, when evaluating distinct counterfactuals with the same antecedent on any single occasion, even when the context underlying the evaluation of each counterfactual may vary. Alternative approaches go some of the way toward resolving this inadequacy by appealing to a difference in the consequents associated with counterfactuals with the same antecedent. That is, in addition to the world of evaluation and the antecedent, the *consequent* contributes to the counterfactual's evaluation. But these alternative approaches nevertheless give a single, determinate truth value to any single conditional (same antecedent and consequent), despite the possibility that this value may vary with context. My reply to these shortcomings (chapter 4) takes the form of an analysis of a language that makes appropriate explicit access to the intended context available. That is, I give an account of a contextualized counterfactual of the form 'In context C: *If it were the case that … , then it would be the case that …*'. Although my proposal is largely based on Lewis' (1973, 1981) analyses of counterfactuals (the logic **VW** and its ordering semantics), it does not require that any particular logic of counterfactuals should serve as its basis – rather, it is a general prescription for contextualizing a conditional language. The advantage of working with ordering semantics stems from existing results (which I apply and develop) concerning the properties of ordering frames that facilitate fashioning and implementing a notion of contextual information preservation.

Analyses of counterfactuals, such as Lewis' (1973), that cash out the truth of counterfactuals in terms of the corresponding material conditional's truth at possible worlds result in all counterpossibles being evaluated as vacuously true. This is because antecedents of counterpossibles are not true at any possible world, by definition. Such *vacuist* analyses have already been identified and challenged by a number of authors. I join this critical front, and drawing on existing proposals, I develop an impossible world semantics for a *non-vacuist* account of counterpossibles (chapter 5), by modifying the same system and semantics that serve the basis of the contextualized account offered in chapter 4, i.e. Lewis' (1986) ordering semantics for the logic **VW**. I critically evaluate the advantages and disadvantages of key conditions on the ordering of worlds on the extended domain and show that

there is a sense in which all of Lewis' analysis of mere counterfactuals can be preserved, whilst offering an analysis of counterpossibles that meets our intuitions.

The first part of chapter 1 consists of an outline of the usefulness of impossible worlds across philosophical analyses and logic. That outline in conjunction with a critical evaluation of Lewis' logical arguments in favour of vacuism in chapter 2, and his marvellous mountain argument against impossible worlds in chapter 3, serves to motivate and justify the impossible world semantics for counterpossibles proposed in chapter 5. The second part of chapter 1 discusses the limitations that various conditional logics face when tasked to give an adequate treatment of the influence of context. That introductory discussion in conjunction with an overview of conditional logics and their various semantics in chapter 2 – which includes an in-depth exposition of Stalnaker-Lewis similarity semantics for counterfactuals – serves as the motivation and conceptual basis for the contextualized account of counterfactuals proposed in chapter 4.

Declaration by author

This thesis *is composed of my original work, and contains* no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, financial support and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my higher degree by research candidature and does not include a substantial part of work that has been submitted *to qualify for the award of any* other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the policy and procedures of The University of Queensland, the thesis be made available for research and study in accordance with the Copyright Act 1968 unless a period of embargo has been approved by the Dean of the Graduate School.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis and have sought permission from co-authors for any jointly authored works included in the thesis.

Publications included in this thesis

No publications included.

Submitted manuscripts included in this thesis

No manuscripts submitted for publication.

Other publications during candidature

Popieluch, Mariusz (2016). *Toward a Similarity Semantics for Counterlogicals.* PhDs in Logic VIII, Darmstadt, Germany, 9-11 March 2016.

Contributions by others to the thesis

No contributions by others.

Statement of parts of the thesis submitted to qualify for the award of another degree

No works submitted towards another degree have been included in this thesis.

Research Involving Human or Animal Subjects

No animal or human subjects were involved in this research.

Acknowledgements

I would like to thank my supervisors. First and foremost, I would like to express sincere gratitude to my supervisor Dominic Hyde for his invaluable critical input and intellectual inspiration, boundless patience, encouragement and guidance. Without his wholehearted and unfailing support over the years, which by far surpassed any formal responsibilities or expectations, this thesis would not have been completed. I am grateful to my supervisor Toby Meadows for his meticulous and insightful critical input in ensuring that key parts of the thesis meet the highest of standards. The thesis has improved immensely as a direct consequence of working with him. I would also like to thank Joel Katzav for his diligent reading of the entire thesis draft, valuable feedback during the final stages of the writing process, and guidance in ensuring that the thesis is completed on time.

I would also like to thank the examination committee – Franz Berto and Ed Mares – for their valuable critical remarks and advice regarding publication of the dissertation's content.

Financial support

This research was supported by Australian Postgraduate Award Scholarship.

Keywords

ordering semantics, counterfactuals, counterpossibles, non-vacuism, impossible worlds, comparative similarity, context

Australian and New Zealand Standard Research Classifications (ANZSRC)

ANZSRC code: 220308, Logic, 60%

ANZSRC code: 220309, Metaphysics, 10%

ANZSRC code: 220313, Philosophy of Language, 20%

ANZSRC code: 010107, Mathematical Logic, 10%

Fields of Research (FoR) Classification

FoR code: 2203, Philosophy, 90%

FoR code: 0101, Pure Mathematics, 10%

Contents

Chapter 1

The role of Impossible Worlds in Philosophical Analysis, and The Problem of Context in the Analysis of Counterfactuals

To retain all the new techniques of algebra that brought in not only 'minus' quantities but also their square roots, and to escape the 'impossible' status of the last.

Gerolamo Cardano, *De aliza*, 1570.

And whether two actions are instances of the same behavior depends upon how we take them; a response to the command, "Do that again", may well be the question: "Do what again? Swat another fly or move choreographically the same way?"

Goodman (1972)

But there may be a relevant difference in the occasions of evaluation, even when both the antecedent and the consequent of the conditional remain the same.

Nute (1980)

1.1 Introduction

This chapter diagnoses the symptoms of analytic inadequacy evident in various formal accounts of counterfactual conditionals concerning two independent aspects – treatment of context and of impossible antecedents – and subsequently outlines the corresponding remedies. The first part of the chapter motivates impossible world semantics for counterpossibles by giving an introductory overview of the idea of impossible worlds and its success in applications to philosophical analysis, and the second part contains an introduction to the context-related issues that burden the analysis of counterfactuals (and subjunctive conditionals in general) and motivates an approach that is developed in the thesis.

The introductory overview of impossible worlds begins with intuitive and preliminary characterizations (§1.2), by noting their conceptual kinship to possible worlds and presenting their standard definitions and classifications. After this general introduction I give a detailed survey of a selection of notable applications (§1.3) of impossible worlds in philosophical analysis and logic. Throughout this survey emphasis is placed on how the idea and character of impossible worlds is a natural generalization of its conceptual and historical predecessor – the idea of possible worlds, by demonstrating how impossible worlds fare in their roles of aiding and extending possible world semantics whenever it proves inadequate. The selected applications, which I present in detail, are impossible world semantics for non-normal modal logics (§1.3.1.1) and relevant logics (§1.3.1.2).

After an informal and general overview of possible world semantics for propositional content in terms of intensions (§1.3.2.1), I describe the granularity problem (§1.3.2.2) and how impossible worlds aid the analysis of hyperintensions, and counterpossibles (§1.3.2.3). The only formal application to hyperintensionality that I present is an impossible world semantics for epistemic and doxastic logics that avoids the pitfalls of omniscience and omnidoxasticity, and does so by helping to fashion more realistic (non-ideal) models of the corresponding propositional attitudes of knowledge and belief (§1.3.3.1). The inadequacy of Lewis' (1973) account of counterpossibles, which treats all counterpossibles as vacuously true, is a direct consequence of his rejection of impossible worlds. I give a critical evaluation of his reasons for doing so in chapters 2 and 3. This inadequacy is amended in the form of an account of counterpossibles and its impossible world semantics, given in chapter 5.

The second part of the chapter (§1.4) is devoted to outlining the notoriously persistent, context-related issues that burden the analysis of subjunctive conditionals – in particular, the analysis of counterfactuals. I present and discuss (§1.4.1) famous examples from Goodman (1954) and Quine (1966) and evaluate the advantages (§1.4.3) and limitations (§1.4.4) of Gabbay's (1972) response, which goes some way to resolving the context related issues that the examples illustrate. This evaluation is carried out by comparing Gabbay's account to other notable approaches to counterfactual analysis. To facilitate this comparison, I formalize the pertinent and unique features of Gabbay's account (§1.4.2) in terms of frames that employ Priest's (2018) notion of *imported information*. The entire discussion in this second part of the chapter serves to motivate the account of contextualized (context relativized) counterfactuals proposed in chapter 4, which accommodates and accounts for the changes in

a counterfactual's truth value contingent on the context of its use. I close this discussion and the entire chapter (§1.4.5-1.4.6), by indicating how the account given in chapter 4 – whilst largely drawing on Lewis' (1973, 1981) proposals – develops suggestions from Berto (2014, 2017) and Nolan (1997).

1.2 Preliminaries

1.2.1 What are impossible worlds?

Most of us will agree that the world may have turned out in ways other than it actually has. Some facts may not have been the case, while some other states of affairs may have, contrary to fact, come to be the case.[1] Reference to such ways is ubiquitous in everyday language, e.g. "What would have been the consequences if the NASA *Curiosity* rover had, contrary to fact, found evidence of life on Mars during its mission?" Perhaps even the fundamental laws of nature may have turned out other than they actually are. Similarly, it seems right to say that the world could not just have turned out *any* old way. That is, most would also agree that there are ways in which the world just *could not* have turned out. But more pertinently, what are the criteria for ways that the world couldn't have turned out? Insofar as we can aid our talk and understanding of 'the ways that the world may have otherwise turned out' with the idea of *possible worlds*, it seems natural to aid our talk and understanding of 'the ways that the world couldn't have turned out' with the idea of *impossible worlds*.[2] Or in short, just as we think of possible worlds as representing ways in which the world could be, we can think of impossible worlds as ways in which the world cannot be. But before such an idea can adequately serve as a means to aid our understanding, we should say what we mean by *impossible* and *impossible world*.

1.2.2 Logical impossibility

Impossible worlds then can serve as a paraphrase of the ways the world *could not* have turned out, much in the same way as possible worlds serve as a paraphrase of the ways that the

[1] Note that no ontological (metaphysical) qualification of any kind is assumed of 'ways', and I shall not make any commitments in that regard throughout the thesis.

[2] This is an extended version of an argument given by Lewis (1973, p.84). The original argument has been coined the argument from 'ways', or from 'admissible paraphrase'. The extended version of the argument from ways, i.e. in support of impossible worlds (which can also be read as a *reductio* argument against them), was first formulated by (Naylor 1986, p.29). See (Yagisawa 1988, p.183) for a modal realist account, (Vander Laan 1997, p.598) for an abstractionist account, and (Berto 2009, p.3) for a hybrid account, whereby "possible worlds are taken as concrete Lewisian worlds, and impossibilities are represented as set-theoretic constructions out of them".

world *could* have turned out. The criterion for impossibility however requires clarification. After all, there are different kinds of impossibility – historical, deontic, physical, epistemic, logical, mathematical, and metaphysical. The following thesis, for most part, focuses on *logical impossibility*, and accordingly – *logically impossible worlds*. Focus on logical impossibility narrows down the meaning of 'impossible world' substantially, but that refinement still has a number of meanings in the literature. For example, a 'logically impossible world' may either mean, a world where (i) the logically impossible happens, or merely that (ii) the laws of logic are different. The two are not the same – (ii) doesn't entail (i), because even if the laws were different, they need not manifest themselves. Consider Priest's (2008, p.172) intuitive analogy illustrating this point:

> Note that one might take 'logically impossible world' to mean something other than 'world where the laws of logic are different'. One might equally take it to mean 'world where the logically impossible happens'. This need not be the same thing. If this is not clear, just consider physically impossible worlds. The fact that the laws of physics are different does not necessarily mean that physically impossible things happen there (though the converse is true). For example, even if the laws of physics were to permit things to accelerate past the speed of light, it does not follow that anything actually would. Things at that world might be accelerating very slowly, and the world might not last long enough for any of them to reach super-luminal speeds.

1.2.3 Logically impossible worlds

Let us spell out the various characterizations of *logically impossible worlds* that an overview of the existing relevant literature reveals. I will proceed from the most specific to most general characterizations. Perhaps the most obvious, and quite restrictive characterization of an impossible world, is one which allows explicitly contradictory states of affairs, i.e. an impossible world is one where pairs of contradictory sentences of the form A and $\sim A$ hold (Lycan 1994). Such a characterization would certainly apply if we are considering classical logic, since contradictory pairs are not satisfied by any classical interpretation.[3] Thus, such a world is classically logically impossible. Continuing in this manner, a more general definition

[3] This is an introductory overview, where finer distinctions of what contradictions are, is omitted. For there may be worlds, and reasoning systems, tolerant of contradictory states of affairs, e.g. whereby some A holds and ~A holds the conjunction of A and ~A holding. For an explicit non-adjunctive account see Varzi (1997), and for *schematic* nonstandard worlds that display that property see Rescher and Brandom (1980).

of a (classically) logically impossible world would be achieved by lifting the restriction from *pairs* of sentences, to including inconsistent *sets* of sentences, i.e. sets of sentences containing contradictory pairs. That is, a (classically) logically impossible world would be one whose set of things holding at it is not satisfied by any (classical) interpretation (Priest 1997). Moreover, worlds may be logically impossible not merely by virtue of certain things *holding* at them, but also of certain things *failing to hold* at them that nevertheless hold for all classical interpretations – just consider some world where $A \vee \sim A$ fails to hold for some sentence A.[4]

But obviously failure of things holding at a world to be satisfied by any (classical) interpretation need not imply that it is not satisfied by *any* non-classical interpretation. That is, a logically impossible world may be one governed by a logic other than classical logic. Then we would say that it is a world where the laws of logic are different. A natural way to generalize such a characterization of logically impossible worlds is to refrain from assuming some particular logic to be true (correct), and instead relativize logical impossibility to arbitrary logic L.[5] Then by analogy to classically impossible worlds, an L-impossible world would just be one where the set of things that hold at it do not hold for any L-interpretation. We may for instance deem some paraconsistent logic L as the correct logic – then worlds where the laws of L fail would be logically impossible from the perspective of L. We look at such worlds in the context of relevant logics, in §1.3.1.2. Finally, we may consider worlds that fail to satisfy *any* logical closures.

1.2.4 Closed worlds: deductive closure, closure under entailment

Priest (1997, 2005) refers to the classes of worlds just described as logically impossible worlds – just as we can think of physically impossible worlds as those where the laws of logic are different from the ones in the actual world, we think of logically impossible worlds as those worlds where the logical laws are different than those that actually obtain.[6] In this sense 'logically impossible worlds' is taken to mean 'worlds where the laws of logic are other than those of the logic that is thought to govern the actual world'. Another way of

[4] Cresswell (1970, p.354) describes such worlds, dubbing them 'non-classical'. There, the author employs such worlds in giving a semantics for a family of hyperintensional logics, which he calls 'weakly intensional' logics.
[5] A logical pluralist may even refrain from that assumption, i.e. there needn't be any single correct logic.
[6] For an overview of logically impossible worlds see Priest (1997, pp.401-2), Priest (2001, Ch. 9), and Priest (2005, Ch.1). Also see the latter for more on the open/closed world distinction. See Nolan (1997, p.542) for his unrestricted comprehension principle for impossible worlds.

characterizing such worlds, is to say that such worlds are *closed* under entailment. Informally, this would mean that if a sentence φ (or any formula expressing some proposition) holds at some world w and φ logically entails ψ (for some logic L) then ψ also holds at w. Since such worlds are closed under entailment, Priest (2005) has coined this class *closed worlds*.[7]

1.2.5 Open worlds

There is another broad class of impossible worlds that violate any kind of closure, save for identity (trivial consequence), i.e. save for the principle: if A holds at world w, then A holds at world w. I follow Priest's (2005) terminology and refer to them as *open worlds*.[8]
Specification of truth of open worlds is analogous to that of logically impossible (closed) worlds. Whereas, at logically impossible, closed worlds, modal (intensional) formulae are treated as atomic in terms of truth value assignment in any given model, at open worlds *all* formulae are treated this way. This means that at open worlds even the truth values of extensional formulae, i.e. containing only extensional connectives, do not respect recursive specification. For example, given some open world where $p \wedge q$ is true, p need not be true there, or $\neg\neg p$ may be true without p being true. Even $A \wedge B$ and $B \wedge A$ need not have the same values at open worlds. This is how Priest motivates open worlds in the context of intensionality:

> Just as there are worlds that realize the way that things are conceived to be when that conception is logically possible, and worlds that realize how things are conceived to be when that conception is logically impossible, so there must be worlds that realize how things are conceived to be for the contents of arbitrary intentional states. Since such states are not closed under entailment, neither are these worlds. We are therefore led to posit a class of unclosed, or *open*, worlds. (Priest, 2005, pp.21-2)

A way of characterizing a class of worlds that would *include* open worlds, is to place no restriction on the reading of the phrase 'ways things could not have been'. A notable example of such a general characterization of impossible worlds is due to Nolan:

---

[7] Deductive closure of a set of sentences is to be understood as closure under logical consequence when the rules of deduction in question are logical.
[8] Some authors

> I think the most plausible comprehension principle for impossible worlds is that for every proposition which cannot be true, there is an impossible world where that proposition is true. This comprehension principle, while natural, will be inconsistent with most accounts of impossible worlds, according to which impossible worlds obey some constraints, but not as many as possible worlds. [It] is at least a good working hypothesis […] we do not, it seems to me, require that the specifications of ways things cannot happen meet any particular requirement, except that they not be ways things could happen. (Nolan 1997, p.542)

That his comprehension principle includes open worlds Nolan makes explicit in the model theory, by lifting any constrains on truth-value assignments to formulae at impossible worlds.[9]

> As I say on p. 542, I think that a very generous comprehension principle for impossibilities is called for: and I model this by not putting any constraints on assignment of truth values to propositions at impossible worlds. (*Ibid*, p.562)

Such a characterization includes all impossible worlds conceivable – historical, nomical (physical), metaphysical, mathematical or logical kind, and any that I have not mentioned here. It's an open-ended definition, pertaining to *any ways* that just *could not be*, and as such is useful to delineate what we mean when we speak of impossible worlds.

Open worlds have proved to be helpful in in the analysis of intensionality, e.g. fashioning epistemic and doxastic logics that avoid counterintuitive consequences such as omniscience and omnidoxasticity. Historically, this particular context of application was also where the characterization of open worlds first appears. [10] (I attend to those issues in more detail in §1.3.2 and to how open worlds akin to open worlds resolve them in §1.3.4.1).

1.3 Why impossible worlds?

Impossible worlds possess a proudly robust track record of successfully extending the role

[9] To see that Nolan's (1997) comprehension principle – modelled in this manner – includes open worlds, let us just consider any closure principle (valid inference form) C: if all premises $A_1, A_2, \ldots$ are true, then the conclusion B is true. Nolan's model theory – which I take to be an attempt to make his comprehension principle formally precise – permits the existence of a world where all of $A_1, A_1, \ldots$ are true but B is not true. Now since C was an arbitrary closure principle, Nolan's comprehension principle includes open worlds.

[10] Priest (2005) relies on open worlds in support of the formal aspects of his general proposal for the logic and metaphysics of intentionality.

that possible worlds had played in aiding analyses across a variety of contexts, such as modality, intentionality, counterfactuals, and relevance. In particular, impossible worlds appear to be the natural candidates in at least two important contexts that are yet to see a thorough and adequate treatment. The first of those being the development of an adequate (non-vacuous) theory of *counterpossible reasoning*, and the second being an account of *propositional content* and an adequate analysis of *hyperintensional phenomena*.

One of the major intellectual breakthroughs in early second half of 20th century analytic philosophy that fed off the widespread and fervent shift to analyses of intensionality at the time, was the pioneering work done in possible worlds semantics. This revolutionary shift received a significant impetus by Saul Kripke (1959), who gave a very intuitive interpretation of C.I. Lewis' axiomatic systems S4 and S5. The Kripke, relational semantics approach, had soon found a number of generalizations and applications. Notable, pioneering contributors to this revolution were Dana Scott (1970) and Richard Montague (1970) who jointly developed neighborhood semantics; Robert Stalnaker (1968) and David Lewis (1973) who extended this approach to pioneer and develop a semantics for conditional logics based on world similarity modelled as nearness within stratified neighborhood frames, commonly known as similarity spheres; building on the pioneering insights of H.G. von Wright, Jaakko Hintikka's (1962) interpretation of doxastic and epistemic operators on structures that bear a strong resemblance to Kripke models, consisting of possible states of affairs and an agent's doxastic/epistemic alternatives.

Possible world analysis, along Kripkean lines, had soon found its limits posed by the challenge of giving a semantics for of C.I. Lewis' axiomatic systems weaker than S4. That is, systems where the *Axiom of Necessitation* fails, which expressed in English states that *'all theorems of logic are necessary',* i.e. if $\vdash A$ then $\vdash \Box A$. The corresponding semantic principle, sometimes called the *Rule of Necessitation* can be stated, as: if $\vDash A$ then $\vDash \Box A$.[11] In order to widen the scope of applicability in the same vein, Kripke (1965) introduced impossible worlds (coining them *non-normal*) to his format of analysis. This pioneering technical 'trick' sufficed to give a (sound and complete) semantics with respect to C.I. Lewis' systems weaker than S4, i.e. notably S3 and S2. Kripke's motivation for delivering a semantics for those weaker logics may not have been entirely independent of the fact that C.I. Lewis endorsed S2

[11] Priest (2008, p.68).

as giving the correct account of logical necessity.[12]

1.3.1 Applications of closed worlds: modal logic and relevant logic

1.3.1.1 Kripke semantics for S2 and S3.

Below I give a quick revision of Kripke model theory, which includes possible/impossible world semantics for normal and non-normal systems . It can be skipped by those familiar with the material. First let us start with the basic ingredients for our language $\mathcal{L}$ i.e. a set of propositional variables $PV = \{p_n : n \in \mathbb{N}\}$ the elements of which shall be denoted with lowercase Roman letters $(p, q, r, \dots)$ or subscripted lowercase Roman p's $(p_1, p_2, \dots, p_k, \dots)$, or lowercase Greek letters $(\varphi, \psi, \chi, \dots)$; unary connectives: ~ (negation), □ (necessity), ◊ (possibility); and binary connectives: ∧ (conjunction), ∨ (disjunction), ⊃ (material conditional). For the metalanguage, upper case letters $(A, B, C, \dots)$ shall be used as variables ranging over complex formulae and propositional variables.

Definition 1.0: Define the basic modal language, denoted $\mathcal{L}$, to be the set: $\{\sim, \Box, \Diamond, \wedge, \vee, \supset\}$.

Definition 1.0.1: Let For be the smallest set closed under the following well-formed formula formation rules:

B: All propositional variables are wffs, i.e. $PV \subseteq For$.

R1: If $A \in For$ then $\{\sim A, \Box A, \Diamond A\} \subseteq For$.

R2: If $\{A, B\} \subseteq For$ then $\{A \wedge B, A \vee B, A \supset B, A > B\} \subseteq For$.

Definition 1.1: A *Kripke frame* is a pair (W, R), where W is a set, and $R \subseteq W \times W$.[13] Formally, W is an arbitrary set of objects. On the intended interpretation, relevant to the semantics under consideration, its elements are *possible worlds*. R is called the *accessibility relation*.[14] So aRb is read as 'b is accessible from a', or 'a accesses b'.

[12] (*Ibid*, p.65).

[13] Various constraints on *R*, e.g. reflexivity, symmetry, or transitivity yield a variety of different conceptions of logical necessity. However, I shall not focus here on that category of constraints, and treat *R* in the most general sense, as not to distract from the focus of the discussion, which is the influence of admitting non-normal worlds on Kripke models.

[14] In general, *W* is any set, and its elements, depending on the context of application, have various names, e.g. *points*, *states*, *nodes*, *scenarios*, *worlds*.

Definition 1.2: A *Kripke model* is a triple (W, R, v), where (W, R) is a *Kripke frame*, and for each $w \in W$, $v_w: PV \longrightarrow \{0,1\}$ is the function assigning at each world w either a 0 or 1 to each propositional variable p. Informally we think of $v_w(p) = 1$ as p being true at w in the model and $v_w(p) = 0$ as p being false at w in the model.

Truth in a model is defined in terms the satisfiability relation $\Vdash \subseteq W \times For$. We read $w \Vdash A$ as 'A is true at w'. Given a Kripke model (W, R, V) and any $w \in W$, define $\Vdash$ as follows:

(1) $w \Vdash p$ iff $v_w(p) = 1$

(2) $w \Vdash \sim A$ iff not $w \Vdash A$

(3) $w \Vdash A \wedge B$ iff $w \Vdash A$ and $w \Vdash B$

(4) $w \Vdash A \vee B$ iff $w \Vdash A$ or $w \Vdash B$

(5) $w \Vdash A \supset B$ iff $w \Vdash \sim A$ or $w \Vdash B$[15]

(6) $w \Vdash \Box A$ iff $\forall u \in W$, such that wRu, $u \Vdash A$.

(7) $w \Vdash \Diamond A$ iff $\exists u \in W$, such that wRu, $u \Vdash A$.

When we want to explicitly refer to truth at a world in a particular model $\mathfrak{A}$, we shall employ the following notation: $\mathfrak{A}, w \Vdash A$.

Definition 1.3: K-validity

Let $\vDash_K \subseteq \wp(For) \times For$, and define $\Sigma \vDash_K A$ iff for all Kripke models (W, R, V), and all $w \in W$, if $w \Vdash B$ for all $B \in \Sigma$, then $w \Vdash A$. That is, valid inference is defined as truth preservation at all worlds in all Kripke models. A formula $A \in For$ is said to be valid iff $\emptyset \vDash_K A$.

Now that we've defined Kripke models, and K-validity we can easily define Kripke frames and models that admit non-normal worlds, and the resulting logic. In non-normal models, modal formulae are assigned *fixed* values at non-normal worlds. All *box*-prefixed formulae are assigned the value corresponding to *falsity* in the interpretation, and all the *diamond*-prefixed formulae are assigned the value corresponding to *truth* in the interpretation. The truth conditions for the remaining formulae at non-normal worlds remain unchanged, in particular, propositional tautologies remain true at all worlds.

Definition 1.4: A *Kripke non-normal frame* is a triple (W, N, R), where W and R are as before, but now $N \subseteq W$ is the distinguished set of normal (possible) worlds and $W \backslash N$ is interpreted as the set of non-normal (impossible worlds).

[15] Note that $w \Vdash A \supset B$ iff $w \Vdash \sim A \vee B$ for all $w \in W$.

Definition 1.5: A *Kripke non-normal model* is a quadruple (W, N, R, v), where (W, N, R) is a *Kripke (non-normal worlds) frame*, and v is as on normal Kripke models, with the exception of non-normal worlds, where all formulae with $\Box$ or $\Diamond$ as their main operator (i.e. $\Box$-formulae, and $\Diamond$-formulae), are assigned values directly, i.e. for all $w \in W \backslash N$, and all $A \in For$:

(N6) $v_w(\Box A) = 0$
(N7) $v_w(\Diamond A) = 1$

The only change to validity condition is that formula validity and valid inference are defined as truth at all normal worlds in all models and truth preservation at all normal worlds in all models, respectively. The motivation for this definition of logical truth and validity is justified if we characterize impossible worlds to be those where the laws of logic are different or where the laws of logic fail. Then when we define validity and valid inference, i.e. the laws and rules of logic, we should not consider worlds where the laws of logic are different or where they fail.

Definition 1.6: N-validity
Let $\vDash_N \subseteq \wp(For) \times For$, and define $\Sigma \vDash_N A$ iff for all Kripke models (W, N, R, v), and all $w \in N$, if $w \Vdash B$ for all $B \in \Sigma$, then $w \Vdash A$. A formula $A \in For$ is said to be valid iff $\emptyset \vDash_N A$.

Such extended truth conditions and validity conditions suffice to provide a counterexample to the Rule of Necessitation, thereby yielding a semantics for logics *weaker* than S4. To see this, consider the countermodel: let $\mathfrak{N} = (W, N, R, v)$, $W = \{w, i\}$, $N = \{w\}$, $R = \{(w, w), (w, i)\}$, and $v_w(p) = v_i(p) = 1$. Now, the tautology $p \vee \sim p$ is true at *all* worlds, so in particular it is true at all worlds accessible to w. Hence $w \Vdash \Box(p \vee \sim p)$, and consequently $\vDash \Box(p \vee \sim p)$. However, it is not the case that $\vDash \Box\Box(p \vee \sim p)$. To see this, note that $w \nVdash \Box\Box(p \vee \sim p)$ since it's not the case that at all worlds accessible from w the formula $\Box(p \vee \sim p)$ is true, since we have $i \nVdash \Box(p \vee \sim p)$ by (N6). Hence $\nvDash \Box\Box(p \vee \sim p)$, as required.[16]

[16] A similar result, of invalidating the rule of necessitation, could be achieved by models very similar to the non-normal models just discussed, with the simplification of conditions (N6) and (N7) in a way that the valuation (interpretation) function instead of assigning to all $\Box$-formulae and $\Diamond$-formulae a *fixed* value at non-normal worlds, assigns *arbitrary* values (call it condition (N8)). This approach has been implemented by Cresswell (1966) in a semantics for Lemmon's (1957) modal system S0.5 (Priest 2008, §4.4a; Berto & Jago 2019, §4.2). Note that models that satisfy (N6) and (N7) are subsets of the models satisfying the weaker condition (N8). I highlight this method, because we will encounter similar ones, later in this chapter and in chapter 5.

1.3.1.2 Relevant Logics

Another area of logic where impossible worlds have found a natural application is *relevant logic*. The central idea of relevant logic is that the conclusion must be *relevant* to the premises in a valid inference. So the development of such logic has been motivated by giving a more intuitive characterization of deductive inference with the defining feature that forces the premises of an argument to be *really used* in deriving the conclusion.[17] Among well-known relevance-violating inference patterns are the following, commonly referred to as *paradoxes of strict implication*, where strict implication $\prec$ is defined as $A \prec B := \Box(A \supset B)$.[18]

(P.1) $\Box B \vDash A \prec B$

(P.2) $\sim\Diamond A \vDash A \prec B$

In general, relevant logics aim to avoid any *irrelevant* inference patterns $A \vDash B$ (or valid implications $A \rightarrow B$), i.e. where A and B do not have any propositional variables in common. One way of capturing the relevant connection between the antecedent and consequent is as follows:

Definition 1.7: (Priest 2008, §9.7.8) A propositional logic is *relevant* iff whenever $A \rightarrow B$ (where $\rightarrow$ denotes the conditional, i.e. logical implication) is logically valid, A and B have a propositional variable in common.

Clearly both (P.1) and (P.2) fail to satisfy this condition, since in (P.1) the consequent can be any formula and in (P.2) the antecedent can be any formula, i.e. $A \prec B$ is valid if either A is necessarily false or B is necessarily true. For example, we have the following special cases:

(P.1.1) $\vDash A \prec (B \vee \sim B)$

(P.2.1) $\vDash (A \wedge \sim A) \prec B$

It is easy to show that the above are valid even on the weakest non-normal Kripke systems corresponding to weakest modal logics S2 and S3, which we have discussed in the previous section.[19] By way of example, I will discuss a relevant logic N_4 presented by Priest (2008,

[17] Mares (2004, p.3).

[18] Those are just the modal counterparts of analogous inference patterns characteristic of the classical logic material conditional (paradoxes of the material conditional): $B \vDash A \supset B$ and $\sim B \vDash A \supset B$.

[19] First let's prove (P.1). Let $\mathfrak{N} = (W, N, R, v)$ be any Kripke non-normal model, and $w \Vdash \Box B$ for some $w \in N$. This means that $u \Vdash B$ for all $u \in W$, such that wRu. But that means that $u \Vdash A \supset B$ for all $u \in W$, such that wRu, by definition of $\supset$. Hence $w \Vdash \Box(A \supset B)$, by definition of $\Box$. Hence $w \Vdash A \prec B$ by definition of $\prec$, as required. Now to prove (P.2). Let $\mathfrak{N} = (W, N, R, v)$ be any Kripke non-normal model, and $w \Vdash \sim\Diamond A$ for some

§9.4), in order to illustrate how impossible worlds can aid invalidating *any* relevance violating inference or conditional.[20] The impossible worlds that we'll encounter in this case are various classically impossible (truth value glut or gap admitting) closed worlds, and worlds where the laws of logic are different (logically impossible worlds), based on a similar idea to the ones fashioned by Cresswell (1966).[21] However, here the focus of analysis is not the modal notions of possibility and necessity, but relevant implication.

I will start by outlining the underlying paraconsistent (basis) of N_4, which is the logic called *First Degree Entailment* (FDE) – first formulated by Nuel Benlap in his doctoral dissertation.[22] The language of FDE is the propositional part of the modal language given in *Definition 1.0*, i.e. $\mathcal{L}_{FDE} = \{\sim, \wedge, \vee, \supset\}$, where $A \supset B$ is defined as $\sim A \vee B$. I give Dunn's relational semantics.[23]

Definition 1.8: An FDE interpretation ρ is a binary relation between the set of propositional variables PV and truth values, i.e. $\rho \subseteq PV \times \{0,1\}$ is a relation between PV and $\{0,1\}$. We read $p\rho 1$ as 'p relates to 1' and $p\rho 0$ as 'p relates to 0'. Also, I will employ the notation $(p, x) \in \rho$ and $(p, x) \notin \rho$ for '$p\rho x$' and 'it's not the case that $p\rho x$', for $x \in \{0,1\}$. We extend ρ to the entire set of well-formed formulae For as follows:

$A \wedge B\rho 1$	iff	$A\rho 1$ and $B\rho 1$
$A \wedge B\rho 0$	iff	$A\rho 0$ or $B\rho 0$
$A \vee B\rho 1$	iff	$A\rho 1$ or $B\rho 1$
$A \vee B\rho 0$	iff	$A\rho 0$ and $B\rho 0$
$\sim A\rho 1$	iff	$A\rho 0$
$\sim A\rho 0$	iff	$A\rho 1$

$w \in N$. This means that there is no $u \in W$, such that wRu and $u \Vdash A$, i.e. for all $u \in W$ such that wRu, $u \nVdash A$, i.e. $u \Vdash \sim A$. But this means that $u \Vdash A \supset B$ for all $u \in W$, such that wRu, by definition of $\supset$. Hence $w \Vdash \Box(A \supset B)$, by definition of $\Box$. Hence $w \Vdash A \prec B$ by definition of $\prec$, as required.

[20] The logics K_4 and N_4 given by Priest (2008, §8, §9), and discussed in this section are not to be found in any earlier literature. They serve as a satisfactory illustration of impossible world semantics for relevant logics. N_4 should not be confused with Wansing's (2001) logic of constructible negation, by the same name.

[21] See footnote 16.

[22] For a history of FDE see (Omori & Wansing 2017). I base this presentation of FDE and the definition of N_4 on Priest (2008, §8, §9). Note that relevant logics are paraconsistent, since the classically valid rule of inference known as *ex contradictione quodlibet* : $A \wedge \sim A \vDash B$ is demonstrably relevance violating.

[23] Other notable semantics for FDE are many-valued semantics, and *the Routley star*, where negation is treated as an intensional operator (Priest 2008, §8).

Note that there are no restrictions on ρ in place such that $(p,1) \notin \rho$ wold imply $(p,0) \in \rho$. Nor does $(p,1) \in \rho$ imply $(p,0) \notin \rho$. Call a formula A a *truth value glut* whenever it relates to both 1 and 0 on some interpretation ρ, i.e. $A\rho 1$ and $A\rho 0$ (equivalently, $(p,1) \in \rho$ and $(p,0) \in \rho$), and call a formula A a *truth value gap* whenever it relates to neither 1 nor 0 on some interpretation ρ, i.e. neither $A\rho 1$ nor $A\rho 0$ (equivalently, $(p,1) \notin \rho$ and $(p,0) \notin \rho$).[24] Validity and valid inference are defined as truth for all interpretations and truth preservation for all interpretations, respectively.

Definition 1.9: FDE-validity
Let $\vDash_{FDE} \subseteq \wp(For) \times For$, and define $\Sigma \vDash_{FDE} A$ iff for FDE interpretations ρ: if $B\rho 1$ for all $B \in \Sigma$, then $A\rho 1$. A formula $A \in For$ is said to be valid iff $\emptyset \vDash_{FDE} A$.

It should be noted that both $p \vDash_{FDE} q \vee {\sim}q$ and $p \wedge {\sim}p \vDash_{FDE} q$ fail in FDE. Perhaps this should be obvious for two reasons – because both LEM and LNC are invalid in FDE (since truth value gaps and truth value gluts are allowed), and the formulae in the above inference patterns are independent.[25] Also note that $p \wedge {\sim}p \vDash_{FDE} q$ is closely related to ECQ, the failure of which makes FDE a paraconsistent logic.[26]

Proposition 1.0: $p \nvDash_{FDE} q \vee {\sim}q$, $p \wedge {\sim}p \nvDash_{FDE} q$
Proof : First for the counterexample to $p \vDash_{FDE} q \vee {\sim}q$, let ρ be an FDE interpretation such that $(p,1) \in \rho$, $(q,1) \notin \rho$ and $(q,0) \notin \rho$, i.e. q is a truth value *gap*. It follows that $({\sim}q,1) \notin \rho$ and $({\sim}q,0) \notin \rho$, i.e. ${\sim}q$ is also a truth value *gap*. It follows that $(q,1) \notin \rho$ and $({\sim}q,1) \notin \rho$. Therefore, $(q \vee {\sim}q,1) \notin \rho$, as required. □

Now for a counterexample to $p \wedge {\sim}p \vDash_{FDE} q$, let ρ be an FDE interpretation such that $(q,1) \notin \rho$, $(p,1) \in \rho$ and $(p,0) \in \rho$, i.e. p is a truth value *glut*. It follows that $({\sim}p,0) \in \rho$ and $({\sim}p,1) \in \rho$, i.e. ${\sim}p$ is also a truth value *glut*. It follows that $(p,1) \in \rho$ and $({\sim}p,1) \in \rho$. Therefore,

[24] FDE is a sub-logic of classical logic and a number of other notable 3-valued logics, like Priest's logic of paradox LP and Kleene's 3-valued logic K_3. It's easy to show that FDE interpretations of the propositional language, that don't admit truth value gaps are just LP interpretations, and those that don't admit truth value gluts are just K_3 interpretations. FDE interpretations that don't admit either truth value gluts or gaps are just classical interpretations. See (Priest 2008, §7, §8).
[25] LEM and LNC are acronyms of the laws of classical logic: the *law of excluded middle* $\vDash A \vee {\sim}A$ and the *law of non-contradiction* $\vDash {\sim}(A \wedge {\sim}A)$.
[26] I am saying 'related', because the standard form of ECQ (i.e. *ex contradictione quodlibet*) is from contradictory premises, i.e. $p, {\sim}p \vDash q$. But conjunction simplification, i.e. $p \wedge q \vDash_{FDE} p$ and $p \wedge q \vDash_{FDE} q$ is valid in FDE, hence $p \wedge {\sim}p \nvDash_{FDE} q$ implies $p, {\sim}p \nvDash_{FDE} q$. Hence ECQ is invalid in FDE. See (*Ibid*).

$(p \wedge {\sim}p, 1) \in \rho$, as required. □

We can now define a modal logic that models a counterpart to strict implication – one that inherits the relevant features of the base logic FDE. Informally speaking, the idea is to define a new (object language) intensional connective → and endow it with the same properties as the metalinguistic relation $\vDash_{FDE}$, i.e. have $A \rightarrow B$ modeled with $A \vDash_{FDE} B$. For a formal treatment we require a proper model theory. The system I'll describe is what Priest (2008, §9.2-9.3) calls K_4.[27]

The language of K_4 is just the propositional language expanded by intensional connective →, intended to represent the relevant conditional, i.e. $\mathcal{L}_{K_4} = \{{\sim}, \wedge, \vee, \supset, \rightarrow\}$.

Definition 1.10: A K_4 model is a pair $\mathfrak{A} = (W, \rho)$ where W is a non-empty set, regarded as a set of possible worlds, and $\rho = \{\rho_i : i \in W\}$ is a set of world-indexed relations $\rho_i \subseteq PV \times \{0,1\}$, such that each ρ_i is an FDE interpretation.[28] The truth and falsity conditions for the propositional part of $\mathcal{L}_{K_4}$, i.e. extensional connectives, are as given in *Definition 1.8*, only they're relativized to worlds, e.g. $A \wedge B\rho_i 1$ iff $A\rho_i 1$ and $B\rho_i 1$. Each ρ_i is extended to account for the intensional connective → are as follows:

$A \rightarrow B\rho_i 1$	iff	$B\rho_j 1$ for all $j \in W$ such that $B\rho_j 1$
$A \rightarrow B\rho_i 0$	iff	$A\rho_j 1$ and $B\rho_j 0$ for some $j \in W$

Call a world $i \in W$ *gappy* if it contains any formulae that are truth value *gaps*, and *glutty* if it contains any formulae that are truth value *gluts*. Validity and valid inference are defined as truth at all worlds in all models and truth preservation at all worlds in all models, respectively.

[27] The name is an abbreviation for $K\upsilon_4$ where 'K' indicates its Kripke structure and 'υ' (upsilon) indicates that the accessibility relation is an equivalence relation, much like in the Kripke system corresponding to S5 (Priest 2008, §3.5). The subscript '4' indicates that we're dealing with a 4-valued logic. That becomes explicit on many-valued semantics for FDE, but on our Dunn-approach it should be noted that there are 4 ways that truth values can be related to a propositional variable. Denote the image of some $p \in PV$ under ρ with $\rho[p] \coloneqq \{x \in \{0,1\} : p\rho x\}$. Then $\rho[p]$ could be any of the following four images: $\{0\}, \{1\}, \{0,1\}, \emptyset$, i.e. false and false only, true and true only, true and false (a truth value glut), and nether true nor false (a truth value gap).

[28] Since the accessibility relation is universal, i.e. $R = W \times W$, it can be accounted for – as implicit – in the truth and falsity conditions for the intensional connective →.

Definition 1.11: K_4-validity
Let $\vDash_{K_4} \subseteq \wp(For) \times For$, and define $\Sigma \vDash_{K_4} A$ iff for every K_4 model $\mathfrak{A} = (W, \rho)$, and all $i \in W$: if $B\rho_i 1$ for all $B \in \Sigma$, then $A\rho_i 1$. A formula $A \in For$ is said to be valid iff $\emptyset \vDash_{K_4} A$.

Perhaps unsurprisingly → counterparts of (P.1.1) and (P.2.1) fail in K_4, as required. That is:
Corollary 1.1: $\nvDash_{K_4} p \rightarrow (q \vee {\sim}q)$, $\nvDash_{K_4} (p \wedge {\sim}p) \rightarrow q$
Proof : First a countermodel to $\vDash_{K_4} p \rightarrow (q \vee {\sim}q)$. It is really a corollary of *Proposition 1.0*, which proves the existence of an FDE interpretation ρ such that $(p, 1) \in \rho$ and $(q \vee {\sim}q, 1) \notin \rho$. Denote that interpretation with ρ^*. Now let (W, ρ) be a K_4 model such that $W = \{i\}$ and $\rho_i = \rho^*$. Then $(p, 1) \in \rho_i$ and $(q \vee {\sim}q, 1) \notin \rho_i$. Hence $(p \rightarrow (q \vee {\sim}q), 1) \notin \rho_i$, as required. □

Now for a countermodel to $\vDash_{K_4} (p \wedge {\sim}p) \rightarrow q$. Similarly, we use the FDE interpretation ρ from the proof of *Proposition 1.0* such that $(p \wedge {\sim}p, 1) \in \rho$ and $(q, 1) \notin \rho$. Denote that interpretation with $\rho^\dagger$. Now let (W, ρ) be a K_4 model such that $W = \{i\}$ and $\rho_i = \rho^\dagger$. Then $(p \wedge {\sim}p, 1) \in \rho_i$ and $(q, 1) \notin \rho_i$. Hence $\big((p \wedge {\sim}p) \rightarrow q, 1\big) \notin \rho_i$, as required. □

Let us briefly examine the class of worlds just defined and see how they fit the classifications of impossible worlds outlined at the beginning of the chapter. Note that among K_4 worlds are classically impossible worlds, since LNC and LEM fail at them, as we've seen directly in the proofs of *Proposition 1.0* and *Corollary 1.1*. In fact, such worlds are impossible for any logic among whose logical laws are LNC and LEM. However, each K_4-world is not only closed by definition (i.e. K_4-closed) but also closed in a variety of other interesting ways if we consider the propositional part of $\mathcal{L}_{K_4}$. It can be easily shown that non-gappy worlds are LP-closed and non-glutty-worlds are K_3-closed, and worlds that admit nether truth value gaps or gluts are classically-closed.[29]

As we have shown, K_4 does go a fair way toward giving an account of a relevant conditional, but some problems remain. Note that if $\vDash_{K_4} A$ then $\vDash_{K_4} B \rightarrow A$. This is clear, since if $(A, 1) \in \rho_i$ for all K_4 models (W, ρ) and $i \in W$, then in particular $(A, 1) \in \rho_j$ for all $j \in W$ such that $(B, 1) \in \rho_j$. In particular, given that $\vDash_{K_4} p \rightarrow p$, it follows that $\vDash_{K_4} q \rightarrow (p \rightarrow p)$, which is

[29] LP denotes Priest's paraconsistent logic (logic of paradox) and K_3 denotes Kleene's 3-valued logic. It suffices to observe that the propositional part of $\mathcal{L}_{K_4}$ is FDE-closed at each K_4-world, by definition. For the rest of the argument see footnote 25.

demonstrably a relevance-violating formula. Therefore, despite invalidating a number of relevance violating formulae, K_4 is *not* a relevant logic, by *Definition 1.7*.
To avoid such commitments, we can introduce worlds where K_4-valid formulae fail to be true. That is if we take → as expressing the laws of logic, we need worlds where those laws can fail. That is, we need to consider worlds where formulae of the form $A \rightarrow B$ can take values other than they take in K_4. [30] This can be achieved by a similar method as employed by Kripke, that is, by assigning truth values *directly* to □-formulae, and ◊-formulae (formulae with a modal operator as the main connective) at non-normal worlds. However, for our purposes we will adopt the method employed by Cresswell (1966) of directly assigning *arbitrary* truth values to such formulae, since we do not wish to presuppose how different the logical laws are at non-normal worlds, just that they *are* different.

Definition 1.12: An N_4 model is a triple $\mathfrak{A} = (W, N, \rho)$ where W is a non-empty set, regarded as a set of worlds, $N \subseteq W$ is the set of normal (possible) worlds and $W \backslash N$ is interpreted as the set of non-normal (impossible worlds), and for each $i \in N$, ρ_i is a set of world-indexed relations of the form $\rho_i \subseteq PV \times \{0,1\}$, i.e. just like on on K_4 models. The only modification to ρ (of K_4 models) is the additional condition for non-normal worlds. That is, for each $i \in W \backslash N$, ρ_i is a set of world-indexed relations of the form $\rho_i \subseteq \{A \rightarrow B : A, B \in For\} \times \{0,1\}$.

Truth conditions for all the $\mathcal{L}_{K_4}$ connectives are exactly the same as for K_4 models, with the exception – evident in the definition of $\{\rho_i : i \in W \backslash N\}$ – that truth values of → formulae are not determined recursively, but rather directly by ρ. Validity and valid inference are defined as truth at all worlds in all models and truth preservation at all normal worlds in all models, respectively. The thought here is that after all we are interested in what follows from what at worlds where logic is *not* different.[31] The non-normal worlds defined in N_4 models, above, are those that Priest (2008, §9.7.2) calls *logically impossible worlds*.[32]

Definition 1.13: N_4-validity
Let $\vDash_{N_4} \subseteq \wp(For) \times For$, and define $\Sigma \vDash_N A$ iff for every N_4 model $\mathfrak{A} = (W, N, \rho)$, and all $i \in N$: if $B \rho_i 1$ for all $B \in \Sigma$, then $A \rho_i 1$. A formula $A \in For$ is said to be valid iff $\emptyset \vDash_{K_4} A$.

[30] Priest (2008, §9.4.4, §9.4.5).
[31] *Ibid* (§9.4.9).
[32] Such worlds would be suitable to evaluate counterpossible conditionals, and indeed a very similar construction is used for an account of counterpossibles that I give in chapter 5, albeit where the base logic is classical.

Proposition 1.2: $\nvDash_{N_4} p \to (q \to q)$

Proof : Let (W, N, ρ) be an N_4 model where $W = \{i, j\}, N = \{i\}$, and let ρ be such that $(p, 1) \in \rho_j$ and $(q \to q, 1) \notin \rho_j$. Hence $(p \to (q \to q), 1) \notin \rho_i$, as required. □

It is relatively easy to show that N_4 is in fact a relevant logic, according to *Definition 1.7*.[33] This concludes our overview of closed, impossible worlds semantics for modal and relevant logics.

1.3.2 Wider applications of impossible worlds

1.3.2.1 Content as intension, via possible worlds

Because this thesis focuses on the analysis of counterfactuals and counterpossibles, I will not give the full characterization of models for intensions and hyperintensions, but merely highlight the demarcation lines where possible worlds fall short of supplying an adequate analysis of these linguistic phenomena. In order to appreciate the vast scope of applicability of impossible worlds, it will be helpful to give an overview of the role that possible worlds have played in aiding philosophical analysis. Also, it will be instructive to give an intuitive outline of those key insights that underpin possible world analysis of intensionality. By doing so, the method's limits will be emphasized.

The general character of such analyses can be traced back to Carnap's (1947) account of content-as-*intension* via possible worlds.[34] Carnap's ideas were developed independently by Montague, Tichý, and Bressan, who all relied on some form of Kripke or Hintikka semantics (Fitting 2015, p.12). The key, underlying idea was to treat intensions as *functions on worlds*. More precisely, intensions were treated as functions from elements of the analyzed language and worlds, to worlds.[35] So if A is a singular term, its intension $\mathcal{I}_A$ is the function $\mathcal{I}_A: (A \times W) \longrightarrow D$ that picks out, for each possible world w, an element $\mathcal{I}_A(w)$ of w's domain D_w corresponding to A's referent at w. For example, the intension of the *singular term* 'the first man on the Moon' would be the function that picks out the individual at each possible world that happened to be the first man on the Moon. The intension would pick out Neil Armstrong in the actual world, and possibly other Apollo 11 mission crew members at other

[33] For the proof, see (*Ibid*, §9.7.9).

[34] To be precise, Carnap spoke of state descriptions, which were maximally consistent sets of atomic sentences and their negations Fitting (2015).

[35] I give a characterization that aims to get across the general idea in a precise way yet without being technically overbearing. Whenever more technical precision will be required, it will be explicitly called upon.

possible worlds. It just so happens that the singular terms 'Neil Armstrong' and 'the first person on the Moon' are co-referential, since their referents coincide at the actual world. At worlds where humans never go to the moon, or worlds where there are no humans the intension will point to nothing. This approach to content analysis has the additional appeal of being in alignment with the intuition that to understand such an expression doesn't require knowledge of its *actual* referent, e.g. 'the tallest tree'.[36]

The general idea is that if A is a meaningful expression, its *intension* is the function $\mathcal{I}_A: (A \times W) \longrightarrow f(D)$ that pics out A's *extension* at each possible world—namely the class of objects at each world of which the expression is true. So if A is a predicate, then for each possible world w, $\mathcal{I}_A(w)$ is a subset of D_w consisting of objects that have the property expressed by A at w. This naturally generalizes to relations on the elements of the domain. So, if A is an *n*-place relation, $\mathcal{I}_A(w)$ is a subset of D_w^n, i.e. its intension picks out, for each possible world w, a set of *n*-tuples of elements of w's domain that stand in the relation expressed by A.[37]

Finally, if A is a sentence, then for each possible world w, $\mathcal{I}_A$ maps to the extension of A's *truth predicate*, namely the set of possible worlds where A is true, i.e. $\mathcal{I}_A: (W) \longrightarrow \{true, false\}$. For the propositional language we can employ Kripke models to give a more precise characterization of $\mathcal{I}_A$.[38] Namely, given a Kripke model $\mathfrak{A}$, let $\mathcal{I}_A(w) = true$ if and only if $\mathfrak{A}, w \Vdash A$, for each $A \in For$. Then for each model $\mathfrak{A}$, the proposition expressed by A can be identified with the set of possible worlds where A is true in that model, or equivalently, the set of possible worlds at which $\mathcal{I}_A$ maps A to truth, i.e. A's intension is the set $\{w \in W : \mathcal{I}_A(w) = true\}$ whose characteristic function is $\mathcal{I}_A$. Effectively, this method has an overall *extensional* character to content analysis.

Intensions deal equally well in drawing distinctions between contingently coextensive predicates. Consider the example from Quine (1951, p.21): although it may just so happen that the property of 'being a creature with a kidney' is coextensive with 'being a creature with a heart', nevertheless these two properties mean something else – it is physically (or at least

[36] Speaks (2014, §2.1.5).
[37] The intension/extension distinction can be traced back to Frege's sense/reference distinction, where *intension* corresponds to *the expressions' meaning*, and *extension* to the *things the expression designates* (Fitting, 2015).
[38] This formulation can be traced back to Carnap's (1947) early work on intensionality.

logically) possible for there to be creatures with hearts, but no kidneys, and the intensions of the two predicates would simply come apart at other possible worlds (as desired). A *sentence's intension*, understood as *the proposition it expresses*, points to that sentence's truth value at each world. The intended interpretation of the set $\{w \in W : \mathcal{I}_A(w) = true\}$ is *the proposition expressed by A.* Possible world semantics for propositional content as intension, conceived this way, had an enormous impact on philosophical analysis in the second half of the 20th century, before its general limitations began to appear.

1.3.2.2 General limitations: "The Granularity Problem"

However, this approach falls short of giving an adequate analysis of contexts containing hyperintensional phenomena, i.e. contexts where *intensional* equivalence is insufficient for identity, or contexts that do not respect *logical equivalence* (Cresswell, 1975, p.25), or more broadly, contexts that do not respect *necessary equivalence* (Nolan, 2013, p.366).[39] Before demonstrating how intensions fall short of delivering hyperintensional distinctions, I will briefly discuss the analogous phenomenon, of the inadequacy that extensions display in drawing intensional distinctions. This will place the following discussion in a broader context. Consider the following example. 'Canberra is the capital of Australia' – the referent of 'Canberra' and 'The capital of Australia' is the city of Canberra.[40] Now, despite its apparent innocuity, the counterfactual 'If Brisbane were the capital of Australia, then Brisbane would be the capital of Australia' gives rise to an intensional context[41], where substitutivity of co-extensive expressions (and co-referential expressions in particular) isn't guaranteed to be truth preserving. Consider an instance of substituting the co-referring singular terms, in this case 'Canberra' for 'the capital of Australia' in the consequent of (2).

1. If Brisbane were the capital of Australia, then Brisbane would be *the capital of Australia*.
2. If Brisbane were the capital of Australia, then Brisbane would be *Canberra*.

Although the first counterfactual is an instance of counterfactual identity, the truth of the

[39] Nolan's generalizes the definition by lifting the restriction to logical equivalence – necessity need not be just *logical*, e.g. it may also be metaphysical, which need not be the same as logical necessity: 'A position in a sentence is said to be sensitive to hyperintensional differences, if the truth value of the sentence can be altered by replacing the expressions in that position with one that *necessarily* applies to the same things' (Nolan, 2013, p.366).

[40] In this example I am treating co-referential terms as a special case of co-extensive predicates, where *being the capital of some country* is thought of as *having that property*.

[41] It is a shift of context, from the actual to the possible in general.

second one is context dependent. In particular there are contexts where it is false, e.g. when we intend (in the hypothetical scenario) both cities to remain where they *actually* are. Brisbane being the capital of Australia, and both cities remaining where they actually are is a perfectly possible scenario, which the reading of (2) doesn't rule out. The general point I wish to stress is that just as co-reference (extensional equivalence) is inadequate for drawing intensional distinctions ('Canberra' doesn't *mean* 'the capital of Australia'), so intensional equivalence falls short of drawing hyperintensional distinctions. Consequently, just as extensional equivalence is inadequate for identity conditions that guarantee substitutivity *salva veritate* on intensional contexts, intensional equivalence is inadequate for identity conditions that guarantee substitutivity *salva veritate* on hyperintensional contexts. That is, the approach fails to distinguish sentences that are either necessarily true, or necessarily false – the former are identified with the set of all possible worlds, since necessity is modelled as truth at all *possible* worlds, and the latter with *the empty set* since necessary falsehoods are not true at any possible world.[42] Let us denote the proposition expressed by sentence A with $[A]$. That is let us adopt the notation $[A] \coloneqq \{w \in W : \mathcal{I}_A(w) = true\}$ for the reminder of our discussion, to distinguish $[A]$ from the sentence that expresses it. To illustrate this more formally, in terms of the Kripke models this would mean that for any two sentences A and B that express a necessary truth, the following identity holds for all models (W, R, V), $[A] = [B] = W$, and likewise, for any two sentences A and B that express a necessary falsehood, the following identity holds for all models $[A] = [B] = \emptyset$ for all models.

Consequently, propositional content modelled as intension leads to cointensive expressions being analyzed as *meaning* the same thing, which is strongly counterintuitive. Consider the following pairs of sentences, which on the just described possible world, meaning-as-intension, analysis are analyzed as expressing the same proposition:

1. There are no married bachelors.
2. There are infinitely many primes.

3. Some bachelors are married.
4. There are finitely many primes.

[42] There is no general consensus regarding what kind of necessity represents absolute necessity, but there is a tendency of taking logical, mathematical, and metaphysical necessity as close approximations. The issue which of these necessities is more fundamental is also controversial (Berto & Jago, 2019, §1.2). In this thesis I will not make any assumptions in this regard.

Each sentence in the first pair expresses some necessary truth, so each is true in all possible worlds. So, their intensions are identical, and consequently both sentences are analyzed as having the same content, i.e. expressing the same proposition. This doesn't seem right, since they appear to be saying different things – (2) says nothing about marriage or bachelorhood. The latter pair of sentences suffers from the same inadequacy of distinguishing their meaning due to the pair being cointensive. This fundamental shortcoming is inherited in the analysis of propositional attitudes that give rise to hyperintensional contexts, where substitutions of necessary equivalent terms in a sentence need not be truth preserving. Take the following pairs of necessarily true sentences:

1. The axioms of Peano Arithmetic are true.
2. [Any sentence that expresses a theorem of PA] is true.

3. Water is water.
4. Water is H_2O.

And consider the following substitutions of those sentences in sentences expressing doxastic and epistemic propositional attitudes:

5. *Giuseppe believes that* the axioms of Peano Arithmetic are true.
6. *Giuseppe believes that* [any sentence that expresses a theorem of PA] is true.

7. *It is known a priori that* water is water.
8. *It is known a priori that* water is H_2O.

It becomes clear that in such contexts, necessary sentences are not expected to be substitutable *salva veritate*. Surely, Giuseppe Peano believed the truth of his own axioms (that's what it means to be an axiom – a truth that is immediately evident), but it doesn't seem true that he believed all sentences, of arbitrarily complexity, that happen to express a consequence of those axioms, i.e. all theorems of PA. So, although (5) is (very likely) true, sentence (6) should be false. Also, whereas sentence (7) is true (it is probably among the least contested *a priori* truths out there), sentence (8) is false, since knowing that water is H_2O requires empirical knowledge of the molecular structure of water.

Counterpossible reasoning is another context where hyperintensional phenomena arise. A counterpossible conditional is a subjunctive conditional whose antecedent expresses a necessary falsehood. By direct analogy with the earlier example involving counterfactuals,

which illustrated the inadequacy of appeals to *actually* co-referring terms (material equivalence) in drawing intensional distinctions, it can be shown that appeals to *necessarily* co-referring terms (logical equivalence) are inadequate in drawing hyperintensional distinctions. This inadequacy in distinguishing between cointensive impossible expressions carries over to accounts of the counterpossible that restrict the analysis to intensions only.[43] All analyses of counterfactuals whose truth is cashed out in terms of the corresponding material conditional's truth at possible worlds will result in all counterpossibles being evaluated as vacuously true. This is because antecedents of counterpossibles are not true at any possible world, by definition. A notable example of such an approach – one given by Lewis (1973, 1986) – meets the same predicament, by treating *all counterpossibles as logically equivalent* (in the case of Lewis, as *true*). As mentioned earlier, I will discuss Lewis' analysis of counterfactuals and counterpossibles in depth in chapter 2, but for the purposes of the present introduction his analysis can be can be given informally: the counterfactual 'if it were the case that …, then it would be the case that …' is true at a possible world *w* just in case the consequent is true at all the most similar possible worlds to *w* where the antecedent is true. Since sentences expressing a necessary falsehood are true at no possible world, and in particular a possible world satisfying some additional similarity conditions, each counterpossible is analysed as vacuously true. Consider the following pair of counterpossibles. Whereas (1) is clearly true in all contexts, (2) could be false.

1. If Sally were to square the circle, then Sally would have squared the circle.
2. If Sally squared the circle and I doubled the cube, then I would be Sally.[44]

So, on Lewis' analysis (1) and (2) are logically equivalent (both are true at all possible worlds), which seems wrong. For a more emphatic demonstration that counterpossibles do

[43] Observation: (2) appears more readily read as false than its counterfactual analogue of Australia's capital(s), since the property of *being the capital of Australia* counterfactually ascribed to Brisbane – which *actually* belongs to Canberra – is *unique*, whereas in the counterpossible (2) there is explicit talk of *two properties*, which only are identified as meaning the same thing by the underlying (content-as-intension) analysis. Also, independently of such reasons for the apparent disparity in readiness with which we would be inclined to read (2) and the earlier counterfactual example as false, it seems that we tend to "hold on" to numerical identity more than any other properties of objects, i.e. shifts in numerical identity seem to be contextually the most far-fetched. Of course, there are contexts where (2) and its counterfactual analogue would be true, but they don't seem to be among the first ones that we're willing to consider.

[44] Squaring the circle refers to constructing a square of the same area as some circle in a finite number of steps. This construction is mathematically impossible. Doubling the cube is a related, impossible construction, whereby given the edge of a cube one is required to construct the edge of another cube that has twice the volume of the first one.

give rise to hyperintensional contexts, let us consider (3), where we substitute the consequent of (1) with a logically equivalent sentence, and observe that unlike (1), (3) could be false.

3. If Sally were to square the circle, then Mariusz would have doubled the cube.

As a matter of fact, Lewis' analysis of the counterpossibles can be viewed as emblematic of the inadequacy of intensions in drawing hyperintensional distinctions. This example highlights how the matter of non-vacuous counterpossibles and the matter of adequate hyperintensional distinctions are closely related – counterpossibles *do* create hyperintensional contexts.

1.3.2.3 Content as hyperintension, via impossible worlds

A number of philosophers have suggested that one way of meeting the requirements of drawing hyperintensional distinctions, is by admitting impossible worlds to accompany possible ones in our world-semantics for propositional content, and proceeding much in the same general manner as the analysis of intensions on possible worlds.[45] In terms of Kripke structures that would amount to either reinterpreting W as a set of possible *and* impossible worlds, or explicitly adjoining a new set W^* to W (which retains its original interpretation), interpreted as containing impossible worlds, and defining models on the extended domain $W \cup W^*$. This way cointensive expressions would not be represented as coextensive in *all* worlds, since impossible worlds would be precisely where their intensions would come apart, e.g. all necessary truths would remain true in all possible worlds, but some could fail to be true in some impossible worlds, and similarly for necessary falsehoods, as they all would still fail to be true in all possible worlds, but some could be true in some impossible worlds.

> Many of the reasons to switch to a possible-worlds framework for linguistic meaning also support employing a system with impossible worlds as well. Just as a predicate can be coextensive with another without being synonymous, two predicates can have matching extensions in every possible situation and yet fail to be synonymous. (Nolan 2013, p.366)

The proposed refinement to the analysis is still extensional in character, since propositions are identified with subsets of the extended universe $W \cup W^*$, where W is the same as before,

[45] E.g. Hintikka (1975), Rantala (1982), Yagisawa (1988), Priest (2005), Nolan (1997, 2013, 2014), Berto (2010, 2014, 2017), Jago (2009, 2014), Bjerring (2010).

and W^* is a set of impossible worlds. Let's look how the analysis of propositional content on this extended account offers a way of drawing hyperintensional distinctions. Given a sentence A, the function $\mathcal{H}_A: A \times (W \cup W^*) \longrightarrow \wp(W \cup W^*)$, works much in the same way as $\mathcal{I}_A: A \times W \longrightarrow \wp(W)$, with the exception of ranging over the extended universe. But the underlying idea of a sentence's hyperintension $\mathcal{H}_A$ picking out worlds where that sentence is true, remains unchanged, i.e. we identify the hyperintension of A with the set $\{w \in W \cup W^* : \mathcal{H}_A(w) = true\}$, denoted $[\![A]\!]$.
The hyperintensions $\mathcal{H}_A, \mathcal{H}_B$ of cointensive sentences A and B, still agree on their intensions, i.e. $[\![A]\!] \cap W = [\![B]\!] \cap W = W$, but also offer a way for their truth values to come apart at impossible worlds. That is, there may exist an impossible world $w^* \in W^*$ where A holds, but B doesn't, i.e. $w^* \in [\![A]\!]$ but $w^* \notin [\![B]\!]$, which yields the desired result of a semantic distinction between the proposition expressed by A and the proposition expressed by B, i.e. $[\![A]\!] \neq [\![B]\!]$. Hyperintensional distinctions between necessarily falsehoods A and B are achieved much in the same way. Although $[A] = [B] = \emptyset$, it need not be the case that $[\![A]\!] = [\![B]\!] = \emptyset$, for there may exist an impossible world $w^* \in W^*$ where A holds, but B does not hold, i.e. $w^* \in [\![A]\!]$ but $w^* \notin [\![B]\!]$.

The above general discussion has given a general, and largely informal outline of the idea of using impossible worlds in analyzing hyperintensionality. In the next section we look at a particular application of open impossible worlds to hyperintensional propositional attitudes.

1.3.3 Applications of open worlds: doxastic logic and epistemic logic

1.3.3.1 Fine-graining with Rantala worlds

> It is well known that it seems possible to have a situation in which there are two propositions p and q which are logically equivalent and yet are such that a person may believe the one but not the other. If we regard a proposition as a set of possible worlds then two logically equivalent propositions will be identical, and so if "x believes that' is a genuine sentential functor, the situation described in the opening sentence could not arise. I call this the paradox of hyperintensional contexts. (Cresswell 1975, p.25)

The following application of impossible worlds involves the analysis of belief and knowledge. That is, an analysis of contexts where hyperintensional distinctions arise

naturally, and where failing to give an adequate account of such distinctions can lead to very bizarre (incorrect) consequences. Exploiting the intuitive analogies between epistemic and modal propositional attitudes – knowledge and necessity in particular – had opened the door to allowing epistemic (and doxastic) logic enjoy the same intuitive semantics that modal logic had.

Hintikka (1962) pioneered an intuitive and successful Kripke-style interpretation of epistemic language, one in which epistemic space (set of epistemic alternatives) is identified with logical space (possible worlds) and epistemic operators are interpreted in a way analogous to modal operators of the modal language. Knowledge for an agent a is interpreted as truth at all a's epistemic alternatives, i.e. truth at all worlds epistemically possible for a. Consequently, since only possible worlds are the available epistemic alternatives for any agent, then all logical truths are *epistemically necessary* for any agent on the above interpretation. That is, on this interpretation all agents know all logical truths, which is certainly not the case. Moreover, since among logical truths there are entailments, any agent will know all the logical consequences of what they know. This predicament of *logical omniscience* is a direct outcome of the above interpretation. Belief is analyzed analogously, and is burdened with analogous issues.

That is, Hintikka's (1962) analysis sanctioned the following principles, where $\boldsymbol{K}A$ is read as 'it is known that A' and $\boldsymbol{B}A$ is read as 'it is believed that A':[46]

(C1) If $\boldsymbol{K}A$ and $A \vDash B$, then $\boldsymbol{K}B$ (Closure under entailment)
If A is known, and A entails B, then B is known.

If $\boldsymbol{B}A$ and $A \vDash B$, then $\boldsymbol{B}B$
If A is believed, and A entails B, then B is believed.

(C2) If $\vDash A$, then $\boldsymbol{K}A$ (Knowledge of all valid formulae)
If A is a necessary truth, then A is known.

[46] Hintikka (1962) relativizes knowledge and belief to agents, i.e. K_aA and B_aA read as 'agent a knows A' and 'agent a believes A', respectively. But we can simplify the discussion by depersonalizing the analysis, since that is not where the relevant issues are, i.e. although the epistemic/doxastic accessibility relations R_a may be relativized to agents a, and for any to agents a and b, $R_a \neq R_b$, nevertheless both $R_a(w) \subseteq W$ and $R_b(w) \subseteq W$, where W is a set of possible worlds (a model's domain) and $R_x(w) = \{u \in W: wR_xu\}$, i.e. the image of w under R_x.

If $\vDash A$, then $\boldsymbol{B}A$ (Belief in all valid formulae)

If A is a necessary truth, then A is believed.[47]

Both principles do not seem right, as people are neither omniscient, nor do they know all logical consequences of what they know. Similarly, people do not believe all necessary truths or logical consequences of their beliefs. For example, Giuseppe Peano, surely did not know all the theorems of arithmetic (logical consequences of PA axioms), even if he claimed to have justified belief for claiming the truth of PA axioms, i.e. the self-evident nature of their truth. Similarly, he would not believe all conjectures of arithmetic that would turn out to be theorems. The same holds for any other person.

Hintikka's (1975) key insight was to correctly identify the crux of the problem by observing that the theoretical responsibility for logical omniscience was not due to the method of possible world analysis *per se*, but rather the underlying assumption – which he had himself previously endorsed – that "every *epistemically* possible world is *logically* possible".[48] It had been precisely the assumption of such a close analogy between *necessity* and *knowledge* that gave rise to erroneously burdening epistemic logic with logical omniscience. So, if knowledge is not something that is closed under entailment, then perhaps for a more accurate world-analysis of epistemic and doxastic propositional attitudes epistemic and doxastic spaces should be modelled accordingly, by including worlds that fail to be closed under entailment. This is precisely what Hintikka (1975) proposed. Hintikka's suggestion to get around the logical omniscience problem, although retrospectively rather straightforward, marked a revolutionary direction in possible world analysis of propositional attitudes. By abandoning the problematic assumption that all epistemically possible worlds are logically possible, he posited worlds that are *not* logically possible, i.e. "some epistemically possible worlds are not logically possible worlds". [49] The main motivation for adopting impossible worlds as a means to refine the analysis of belief is the now retrospectively obvious observation that human beings are not perfectly (ideal) rational agents:

> The way to solve the problem of logical omniscience is hence to give up the assumption [that every epistemically possible world is logically possible]. This means admitting 'impossible possible worlds', that is, worlds which look possible

[47] Wansing (1990, p.526) Pietarinen (1998, pp.8-9), Berto & Jago (2019, §5.3).
[48] Hintikka (1975, p.476).
[49] Hintikka (1975, p.477).

> and hence must be admissible as epistemic alternatives but which none the less are not logically possible. Admitting them solves our problem for good.
> (Hintikka 1975, p.477)

Within a decade of Kripke's non-normal semantics for S3 and S2, Hintikka (1975) and Rantala (1975) had extended Kripke's model-theoretic "trick" employed in non-normal models, and had developed a semantics for epistemic and doxastic logics that model *non-ideal* agents. Introducing and employing impossible worlds gave a semantics that invalidated epistemic and doxastic versions of problematic closures (C1) and (C2), thereby doing away with omniscience and omnidoxasticity. On the impossible-world semantics this is done by having the valuation function, for each model, assign values directly to formulae at impossible worlds. Effectively, this technical move gives the set of impossible worlds the capacity to violate any closures, including entailment.

For the doxastic logic we expand the propositional language by an epistemic operator symbol $\boldsymbol{B}$, where the intended reading of '$\boldsymbol{B}A$' is 'it is believed that A'. Let the language of basic propositional doxastic logic be $\{\sim, \wedge, \vee, \supset, \boldsymbol{B}\}$. Let $PV = \{p_n : n \in \mathbb{N}\}$ be the set of propositional variables Finally, let For be the smallest set closed under the following formation rules:

B: All propositional variables are wffs, i.e. $PV \subseteq For$.
R1: If $A \in For$ then $\{\sim A, \boldsymbol{B}A\} \subseteq For$.
R2: If $\{A, B\} \subseteq For$ then $\{A \wedge B, A \vee B, A \supset B\} \subseteq For$.

I present simplified Rantala models, which suffice to illustrate the role of impossible worlds in this context, for the present, introductory purposes. Multimodal systems for multiple agents are generally given, where belief is agent-relative and modelled by the corresponding accessibility relation, but for the introductory illustration purposes I only use a single accessibility relation for simplicity. The idea can be easily generalized to accommodate multiple agents.[50]

Definition 1.14: An *Impossible world "Rantala" Model* is the triple (W, W^*, R, v), where W and W^* are nonempty sets, regarded as the set of possible worlds and the set of impossible

[50] For a detailed exposition see Rantala (1982a, 1982b), Wansing (1990), Sillari (2008).

worlds, respectively, and the binary relation $R \subseteq W \cup W^* \times W \cup W^*$ regarded as the accessibility relation.

For each $w \in W$, $v_w: PV \longrightarrow \{0,1\}$ is just as on Kripke models, where the truth conditions for extensional connectives at possible worlds are just like on Kripke models, and:

$v_w(\boldsymbol{B}A)$ iff $\forall u \in W \cup W^*$: if wRu, then $v_u(A)$.

However, the truth or falsity of formulas need not be recursively specified at non-normal worlds. The only restriction is a semantic version of modus ponens, i.e. for all $w^* \in W^*$, and $A, B \in For$:

(†) If $v_w(A) = v_w(A \supset B) = 1$, then $v_w(B) = 1$.

Note that such a constraint precludes Rantala impossible worlds from being fully fledged *open* worlds. The restriction is in place to validate the K-axiom $\boldsymbol{B}(A \supset B) \supset (\boldsymbol{B}A \supset \boldsymbol{B}B)$, which seems appropriate for knowledge and belief. Without it, Rantala impossible worlds would just be open worlds. The main difference between the impossible worlds employed in Kripke non-normal models or N_4 models and the ones employed in Rantala models, is that whereas the former worlds display non-standard behaviour of *intensional* operators only at impossible worlds – box/diamond and → formulae are assigned values directly – in the latter worlds all formulae misbehave this way, which means that even *extensional* connectives behave non-standardly, i.e. they fail to be truth functionally recursive.[51]

Validity and valid inference are defined as truth at all possible worlds, and truth preservation at all possible worlds, respectively. [52]

This way doxastic closure under entailment (C1) fails, since although $A \wedge B \vDash A$, the semantics allows for some impossible world w^*: $v_{w^*}(p \wedge q) = 1$, yet $v_{w^*}(p) = 0$. To see this, suppose that $v_w\big(\boldsymbol{B}(p \wedge q)\big) = 1$ at some possible world w. So, for all worlds $u \in W \cup W^*$ such that wRu, $v_u(p \wedge q) = 1$. Now suppose wRw^*. Hence, we see that $v_w(\boldsymbol{B}p) = 0$, as required. Similarly, such properties of v at impossible worlds are sufficient to violate (C2) allowing belief in any counterexample to any logical law. For example, given $\vDash (A \wedge B) \supset A$, the

[51] So according to Priest's distinction Priest (2005, §1.5), whereas Kripke non-normal worlds are merely *intensionally* impossible, Rantala non-normal worlds, being *extensionally* impossible, display a higher degree of logical deviancy.

[52] Wansing (1990, pp.526-527).

semantics allow for $\boldsymbol{B}\big((p \wedge q) \supset p\big)$ to be invalidated, by choosing a possible world w and impossible world w^* such that wRw^* and letting $v_{w^*}((p \wedge q) \supset p) = 0$, which implies $v_w(\boldsymbol{B}((p \wedge q) \supset p)) = 0$. Hence, $\nvDash \boldsymbol{B}((p \wedge q) \supset p)$, as required. [53]

As we can see, on the open-world approach (practically open if we ignore the (†) rule) described above, whilst the feat of avoiding omnidoxasticity and omniscience can indeed be avoided, it comes at a cost. That is, the proposal goes too far, since it ultimately trivializes belief – anything and everything may be believed, and at the same time any logical or metaphysical truth may be doubted, if no restrictions are place on the impossible worlds. This is a problem. To put it another way, if all open worlds are admitted as legitimate doxastic or epistemic alternatives, then *any* sentence describes a doxastic or epistemic possibility. Conversely, no sentence would be safe from skepticism, even those expressing Frege's cognitively insignificant identities (such as 'water is water'). Surely this isn't right. When distinctions are too fine, we also end up in trouble. The task is to avoid omniscience and omnidoxasticity without ruling out truths that would be self-evident to any rational agent. For proposals that address this problem see Yagisawa (1985), Hawthorne (2005), Chalmers (2010), Jago (2009, 2014), Williamson (2010), and Bjerring (2010, 2012), Berto & Jago (2019).

1.4 Modifying Lewis' account of the counterfactual

1.4.1 Goodman and Quine's *context sensitivity* objections

Counterfactuals are notoriously context sensitive. Take the well-known example:

1. If Caesar had been in command (in Korea), he would have used the atom bomb.
2. If Caesar had been in command (in Korea), he would have used catapults.[54]

Intuitively, the truth of each depends on contextual background assumptions. But how can we tell what they are? For (1) to be true, we require contexts where Caesar's knowledge of modern warfare is assumed to be in line with the military knowledge of a modern military

[53] Examples, and their discussion are borrowed from Pietarinen (1998, p.10), and some definitional layout features from Sillari (2008, p.7).

[54] Quine (1960, p.22) bases this example on similar ones given by Goodman (1954). Lewis (1973, pp.66-7; 1986, p.251) offers a number of replies to the contention pointed out by Quine, but eventually settles on one whereby the uttered counterfactual taken as being asserted, and context is called upon in resolving the vagueness of the comparative similarity in favour of the truth of the uttered counterfactual. This solution seems unsatisfactory since it is ambiguous what context has been called upon in favour of the counterfactual's truth.

general, whereas for the second to be true, no such contextual background assumption is required.

The problem that conditional analyses face is that the role of context is left syntactically ambiguous. That is, at the level of the object language there are no indicators what context should underlie the evaluation of a counterfactual, although Gabbay's (1972) analysis –which we will look at in this section – goes some way toward resolving this ambiguity.

Such context dependence is starkly pronounced in the following pair of counteridenticals given by Goodman (1954). Here the antecedents are logically equivalent (their formulation intends to hint at what the underlying context is), but their consequents are clearly incompatible.[55]

3. If I was Julius Caesar, I wouldn't be alive in the 21th century.
4. If Julius Caesar were I, he would be alive in the 21th century.[56]

It seems that both can be true, or at least they can be heard as true. Indeed, but in different contexts, and they would hardly be true in any single context, which precludes inconsistent scenarios (possibilities). The truth of the above is contingent on what is *meant* by the antecedent, which in turn rests on what context is assumed to underlie the reading of the counterfactual and consequently its evaluation. However, what that context is on any given occasion is not determined by the counterfactual itself. (3) would be true in contexts where the time when Caesar *actually* lived is maintained as true in the hypothetical scenario, whereas for (4) to be true the fact that I am *actually* alive in the 21st century is also true in the hypothetical scenario.

Chapter 2 gives a detailed overview of the semantics, and critical analysis of the general Stalnaker-Lewis similarity account of counterfactual conditionals and how it contrasts with the family of *ceteris paribus* conditionals. For the comparative purposes of this section it will suffice to highlight one fundamental feature that these analyses have in common. On these analyses, when evaluating the truth of a counterfactual at some possible world *w*, only *w* and the *antecedent* determine what set of situations have the features we take to be relevant to our deliberations in evaluating the conditional. In other words, it is only *w* and the *antecedent*

[55] I borrowed this emphasis from Priest (2017, §2.3).
[56] Goodman (1983, p.6).

that determine what context underlies the evaluation of the conditional at w. This stems from the general underlying idea, common to these accounts, of treating the conditional as a special kind of expression of relative necessity.

> We have said that conditionality can be regarded as a species of relative necessity. This idea is bolstered syntactically by redescribing a conditional $A > B$ as $[A]B$ – so that the antecedent A forms a unary operator, the box-like brackets being reminiscent of the operator □ of simple necessity. Thus described, conditionality assumes the aspect of a sententially indexed modality, in which $[A]B$ expresses the necessity of B relative to A. Chellas (1975, p.138)

1.4.2 Gabbay's analysis of subjunctive conditionals

Highlighting that feature of the aforementioned conditional analyses suffices for the task of discussing differences with Gabbay's (1972) analysis of subjunctive conditionals, to which I now turn. Gabbay's analysis follows this general modal idea of analyzing conditionals akin to expressions of sententially indexed modality described by Chellas, but instead of modelling a conditional $A > B$ as $[A]B$ (or equivalently $\Box_A B$), i.e. where the necessity operator is relativized to a single sentential parameter (the antecedent), $A > B$ is modelled as $[A, B](A \supset B)$, following the notation suggested by Chellas, or equivalently $\Box_{A,B}(A \supset B)$. This analysis still uses the same general idea of conditional necessity determining the subset of possible worlds to be considered in the evaluation of the conditional, but it takes its content to be relativized to both the antecedent and the consequent.

Gabbay recognizes the role of context in determining the worlds relevant to the evaluation of the counterfactual, and because – he argues – the consequent carries key contextual information, which the antecedent alone fails to capture, its role is indispensable. In other words, the set of worlds we consider in evaluating $A > B$ is determined by both A and B. More specifically, the consequent determines which contingent aspects (facts) of the actual world are relevant to the evaluation of the conditional, and as such should remain unchanged in the hypothetical scenarios.

> Generally, whenever a statement $A > B$ is uttered at a world t, the speaker has in mind *a certain set of statements* $\Delta(A, B, t)$ (concerning the political situation or geographic situation, etc.) *which is supposed to remain true*, and the speaker wants to express that in all worlds in which all statements of Δ retain their

> truth $A \supset B$ must hold. What is $\Delta(A, B, t)$? Well, one can perhaps find out what Δ is from A, B and the general knowledge and the circumstances at the time of utterance in the world of utterance (i.e. t). The following examples show that Δ depends on both A and B. Consider the statements:
>
> (i) If I were the Pope, I would have allowed the use of the pill in India.
>
> (ii) If I were the Pope, I would have dressed more humbly.
>
> Clearly, in the first statement, we must assume that India remains overpopulated and poor in resources, while in the second example nothing of the sort is required. Gabbay (1972, p.98, emphasis added)

I will adopt Priest's terminology, and refer to whatever carries over invariantly into the relevant hypothetical scenarios (worlds) as information imported from the actual world.

Definition 1.15: Priest (2018, §2.1). Let us call the information that is carried over [from the world of evaluation] the *imported information*.

Applying Priest's terminology to Gabbay's example above, we would say that in the case of (i) we import the information that India is overpopulated, which doesn't seem like the relevant kind of information to import for (ii).

Consider the following pair of counterfactuals, inspired by Goodman, where the contextual input of the consequent is made salient, and which Gabbay uses as one of his examples to demonstrate that Δ depends on both the antecedent and the consequent:[57]

5. If New York were in Georgia, then New York would be in the South.
6. If New York were in Georgia, then Georgia would be in the North.

Gabbay (1972, p.99) points out that at the worlds relevant to the evaluation of (1) and (2) are not the same, despite the counterfactuals sharing the same antecedent. In other words, the information that gets imported into the antecedent worlds for the truth of (5) is not the same as the information imported to the antecedent worlds for the truth of (6). 'Georgia is in the South' must retain its truth value in (5), whereas 'New York is in the North' must retain its truth value in (6).

[57] Gabbay (1972, pp.98-99) gives a similar example.

We can make Gabbay's insight more precise, by an appropriate modification of Kripke frames. First, we add a third recursive clause to *Definition 1.0.1*, to expand the set of wffs For so it contains expressions $A > B$ corresponding to counterfactuals:

R3: If $A, B \in For$ then $A > B \in For$.

We define a world accessibility relation $R_{A,B} \subseteq W \times W$ that depends on both the antecedent and consequent. The worlds in the image of $R_{A,B}$ are regarded as those that are relevant to the evaluation of $A > B$. To give a precise characterization of $R_{A,B}$ in terms of $\Delta(A, B, w)$ let us define models that reflect Gabbay's idea. I am giving an account of Gabbay's analysis in terms of the kind of models (for *ceteris paribus* conditionals) that appear in the next chapter.

Definition 1.16: A *Gabbay frame* is a pair $(W, \{R_{A,B}: (A, B) \in For \times For\})$, where $W \neq \emptyset$, and for each $(A, B) \in For \times For$, $R_{A,B} \subseteq W \times W$ is a *reflexive* relation satisfying:[58]

$$(x, y) \in R_{A,B} \quad \text{iff} \quad y \Vdash C \text{ for all } C \in \Delta(A, B, x)$$

Reflexivity of $R_{A,B}$ is naturally motivated, since $\Delta(A, B, w)$ contains information to be imported, which is already in place at *w* by definition of *imported information*.[59]

Definition 1.16.1: For convenience, define $f_{A,B}(w) := \{u \in W: wR_{A,B}u\}$.

Although Gabbay does not endorse a Stalnaker-Lewis kind of similarity approach to possible world semantics for conditionals, admitting that he feels "uneasy" about that concept, he still allows a similarity-related idiom and calls the worlds in $f_{A,B}(w)$ Δ-similar.[60] This makes his account conceptually closer to the family of *ceteris paribus* conditionals discussed in the next chapter. That is, we can view $f_{A,B}(w)$ as those worlds that are *ceteris paribus* the same as *w*, or at least sharpen the ceteris paribus analysis to include the additional parameter.

[58] What I call *G frames* are special cases of structures widely known among computer scientists as *labelled transition systems* $(W, \{R_a: a \in A\})$, where $W \neq \emptyset$ is a set of states, and $A \neq \emptyset$ is a set of labels (Blackburn 2001, p.3). In the case of Gabbay frames we're clearly labelling the accessibility relation by ordered pairs of elements of *For*.

[59] Gabbay (1972, p.100) emphasizes the reflexivity of the accessibility relation with that exact point, although he doesn't use the information importation terminology.

[60] Gabbay (1972, pp.99-100).

Definition 1.17: A *Gabbay model* is a triple $(\mathfrak{F}, V)$, where $\mathfrak{F}$ is a *Gabbay frame*, $V: PV \longrightarrow \wp(W)$, is the function that assigns to each propositional variable $p \in PV$ a subset of W. Informally we think of $V(p)$ as the set of worlds in the model where is p true.

Truth in a model is defined in terms the satisfiability relation $\Vdash \subseteq W \times For$. We read $w \Vdash A$ as 'A is true at w'. Given a *Gabbay* model $(W, \mathfrak{F}, V)$ and any $w \in W$, define $\Vdash$ as follows:

(1) – (6) are as for Kripke models (see definition 1.2).

(7) $w \Vdash A > B$ iff $\forall u \in W$, such that $wR_{A,B}u$, $u \Vdash A \supset B$.

That is, iff $A \supset B$ is true at all $R_{A,B}$-accessible worlds.

Definition 1.17.1: For convenience, define $[A] := \{w \in W : w \Vdash A\}$.
Equivalently we can express the truth conditions for $A > B$ more concisely in terms of previously defined sets of worlds:

(7') $w \Vdash A > B$ iff $f_{A,B}(w) \cap [A] \subseteq [B]$.

That is, iff B is true at all $R_{A,B}$-accessible worlds.

1.4.3 Advantages of Gabbay's account

Let us highlight some interesting implications of the differences in the treatment of contextual information. From what has already been said, it follows that *ceteris paribus* and Lewis-Stalnaker similarity accounts preclude any context shifts between counterfactuals with the same antecedents in any given situation. Broadly speaking, on any given instance of utterance for counterfactuals with the same antecedent there's a single choice of the worlds that are considered (as relevant) – the underlying context is ambiguous *and* fixed.

Given a pair of counterfactuals with the same antecedents (but different consequents), as in the examples we have looked at, on Gabbay's account, an element of the modelled object language, i.e. the consequent, allows for a context shift on any single occasion, cashed out in terms of considering different sets of worlds when evaluating counterfactuals with different consequents. On *ceteris paribus* and Stalnaker-Lewis similarity accounts there is an ambiguity regarding the context which should underlie the evaluation of counterfactuals in each pair. The inexplicit influence of context reigns at the (metalinguistic) level of models. That is, given any model, a counterfactual's truth value may vary across worlds, but also, crucially, its truth at a single world (say the actual world) varies across models. Gabbay's

analysis provides the additional linguistic parameter in the form of the consequent to decide the matter.

> Gabbay's analysis of conditionals has one apparent advantage that other analyses we've examined lack – it offers a semantic counterpart for the fact, which we have observed, that the evaluation of two conditionals with the same antecedent may require consideration of different sets of situations. (Nute 1980, p.75)

That is, given two conditionals with the same antecedent, Gabbay's analysis can account for a relevant difference in the intended meaning (and evaluation) of the antecedent by appeal to a difference in the consequents associated with the same antecedent. This can be best illustrated by a pair of counterfactuals whose truth depends on such radical context shifts that they can never be jointly true (although they could be both false) on the *ceteris paribus* or Stalnaker-Lewis accounts. Let us revisit the earlier example of the pair of counteridenticals with Caesar.

3. If I was Julius Caesar, I wouldn't be alive in the 21th century.
4. If Julius Caesar were I, he would be alive in the 21th century.

For analyses whose truth conditions depend on the world of evaluation *w* and the antecedent only, as is the case with the *ceteris paribus* and Stalnaker-Lewis analyses, there will be a single set of worlds (the *closest* antecedent worlds to *w*, or antecedent worlds that are *ceteris paribus* the same as *w*) that is considered in the evaluation of both counterfactuals. Consequently, the two are always evaluated as contraries. They cannot be true together because at all those worlds I, Julius Caesar either am or am not alive in the 21st century, in which case either (3) or (4) is true. Alternatively, I, Julius Caesar am alive in the 21st century at some of those worlds and am not alive in the 21st century at others, in which case both (3) and (4) turn out false. There are many more examples of pairs of counterfactuals similar to this one, which I discuss in chapter 4, where I also give formal arguments demonstrating their (contrary) relationship on *ceteris paribus* and Stalnaker-Lewis models.

Let us translate (3) as $A > B$ and (4) as $A > C$. On Gabbay's analysis however, both (3) and (4) can be simultaneously true, since the sets $f_{A,B}(w)$ and $f_{A,C}(w)$ need not be the same, and in particular, they could be disjoint, thereby making it possible for both $f_{A,B}(w) \cap [A] \subseteq [B]$ and $f_{A,C}(w) \cap [A] \subseteq [C]$ to be satisfied. That is $f_{A,B}(w) \cap [A]$ may be the worlds where I (i.e. Julius Caesar) am not alive in the 21st century, and $f_{A,C}(w) \cap [A]$ may be worlds where Julius Caesar

(i.e. me) is alive in the 21st century.

1.4.4 Limitations of Gabbay's analysis

Ignoring the fact that Gabbay's proposed logic invalidates certain very plausible conditional inference forms[61], there's a pertinent shortcoming of particular interest to us, regarding how the analysis fares with drawing finer contextual distinctions.

> The problem is that Gabbay's analysis, just like the other analyses we have examined, will give a single, determinate truth value to the conditional, regardless of the circumstances under which the conditional is evaluated. The formal semantics does not explicitly make provisions for the conditional being accepted on one occasion and rejected on another due to the different circumstances of those occasions. [But] there may be a relevant difference in the occasions of evaluation, […] even when *both* the antecedent and the consequent of the conditional remain the same. (Nute 1980, p.76)

As Nute observes, Gabbay's semantics fares no better than the aforementioned accounts in terms of offering a semantic mechanism that would allow flexibility in reading a conditional (and giving a corresponding truth value) in a manner that accounts for distinct circumstances (intended contextual considerations). Under some circumstances we may read (2) as true and false under others – recall the contextual considerations we discussed earlier.

2. If Caesar had been in command (in Korea), he would have used catapults.

But on the analysis offered by Gabbay, we appear to have run out of syntactic resources present in the conditional that could be employed in making such a distinction. It seems natural to consider a language that would make such explicit access to intended context available, e.g.:

7. In context *a*: If Caesar had been in command (in Korea), he would have used catapults.

If context *a* assumes Caesar's *actual* military knowledge (7) comes out as true, but if context *a* assumes Caesar's military knowledge to be that of a 20th century military general – whilst maintaining his actual traits of a strategic and ruthless genius determined to use the most

[61] See Nute (1980, pp.75-76).

effective means available to him in order to defeat the enemy – then (7) would come out as false.

Note that using the enumeration of conditionals in this section, (7) is just 'in context a: (2)'. The object language, defined in chapter 4, offers a corresponding, context-indexed connective, so for 'in context a: (2)' we would have an expression $A >_a B$ and a different one for 'in context b (2)', i.e. $A >_b B$, both of which need not be evaluated as having the same value. This is the approach that I adopt in the account given in chapter 4.

1.4.5 Berto's context-indexation suggestion

It is difficult to say whether the approach I chose – expanding the language by introducing additional syntactic parameters – to address those concerns is optimal, but it does appear natural. I have found some supporting evidence for this in a recurring suggestion made by Berto (2014, 2017). It was in Berto's work on the analysis of conceivability and imagination that I have found a parallel of what I have been considering in counterfactuals. It was the manner in which Berto (2014, 2017) chose to analyze 'representational acts' underlying our conceivability and imagination, and 'imagination acts' that initially captured my attention, and in particular, his suggestion regarding how one may go about contextualizing those acts. Let me outline those features of Berto's (2017) semantics of imagination that are relevant to his suggestion how one may go about contextualizing the object language.[62]

In Berto's (2017) analysis of imagination, a basic propositional modal language is expanded by the inclusion of a family of sententially indexed modal operators $[A]$, where A ranges over formulae that express possible acts. Expressions central to the analysis $[A]B$, are read as 'It is imagined in act A that B or, more accurately 'It is imagined in the act whose explicit content is A, that B', where B is any well-formed formula.[63] On Berto's analysis $[A]$ acts like a relative necessity operator, ranging over possible and impossible worlds, and $[A]B$ receives an analysis akin to expressions of sententially indexed modality described by Chellas (1975, p.138). Fundamentally Berto's analysis of $[A]B$ rests on the same idea as the one employed in the analysis of *ceteris paribus* conditionals such as $A > B$.

[62] Berto (2014, p.113, f.9) makes the exactly the same suggestion in the context of conceivability.
[63] Berto (2017, §4).

> This brings conceiving in [*sic*] the vicinity of *ceteris paribus* conditionals. The explicit content of a representation may play a role similar to a conditional antecedent. (Berto 2014, p.8)

> The explicit fictional content corresponds to the explicit content of our imagined scenarios, and works, in Lewis' approach, too, like the antecedent of a *ceteris paribus* conditional. (Berto 2017, p.7)

Berto (2014, p.113, f.9; 2017, §5) identifies the same worry in the analysis of intensional states like imagination that had been identified by Quine (1960), concerning the contextual ambiguity of counterfactuals with the same antecedents. "Is it so that, when one imagines in one act $[A]$ that B and that C, one automatically imagines that $B \wedge C$?" – he asks. On a given act of imagination $[A]$ with the same explicit content of Caesar being in command of the US troops in Korea, one can imagine Caesar using the atom bomb B, and one can imagine that he uses catapults C, however it doesn't follow that one would thereby imagine Caesar using both the bomb and catapults. Berto observes that naturally, one *could* also imagine that, but the inference pattern $[A]B, [A]C \vDash [A](B \wedge C)$, should not be an automatic logical entailment.

The heart of the problem rests in the fact that different acts of imagining the same explicit content can give rise to imagining a different scenario in different contexts – in general, the imported information that makes $[A]B$ true is *not the same* as the imported information that makes $[A]C$ true. That is, it seems clear that different contexts underlie the truth of $[A]B$ and $[A]C$, and so it is not obvious that $[A](B \wedge C)$ should follow, *unless we restrict what contexts should be at play throughout the inference*. This can be done via a modification of the object language, by indexing representational acts $[A]$ with contexts, e.g. $[A]_x$, $[A]_y$, which would allow for an explicit syntactic restriction of inferences to a single context, for example $[A]_x\, B, [A]_x\, C \vDash [A]_x\, (B \wedge C)$. Below is how Berto expresses this idea.

> I think that Adjunction can be maintained by *fixing* some contextual parameter. The formalism may represent this, if wanted, by adding a set of contexts to the interpretations and variables ranging on them in the language, and by directly indexing representational acts with contexts: $[A]_x$, $[A]_y$, for instance, will stand for two distinct acts with the same explicit content, A, performed in contexts x

> and *y*. Once the adjunctive inference is parameterized to *same-indexed contents*, it should work fine. Berto (2017, p.11) [64]

Naturally, this solution has its counterpart in the analysis of counterfactuals. Given the already noted similarity between the analysis of Berto's intensional expressions $[A]B$ and *ceteris paribus* (or Lewisean) conditionals expressed as $A > B$, the move to introduce a family $\{>_x : x \in \mathcal{C}\}$ of context-indexed conditional connectives that range over a set of context indices $\mathcal{C}$ seems natural, and it is precisely the method I adopt. One could apply an analogous restriction to the counterpart inference patterns, i.e. $A >_x B, A >_x C \vDash A >_x (B \wedge C)$ in order to ensure truth preservation. In fact I show, in agreement with Berto, that this and other generally accepted inferences[65] hold for the contextualized language, whenever all instances of the counterfactual appearing in the inference are restricted to a single context index. However, I chose to not apply such a restriction in general, which I feel departs from an opportunity to make the logic relevance-sensitive in an interesting way (not only single context index premise sets).

My proposal allows for premises to range over more than one context index, so in this sense the restrictions that are in place are weaker than the one suggested by Berto. However, on top of the usual truth (preservation) condition for validity, a *contextual information preserving* condition is introduced. That is, I demand the existence of relevant content connection (properly defined and developed in the model theory) between the context indices over which the premises range and the conclusion context index. For example, in the case of $A >_x B, A >_y C \vDash A >_z (B \wedge C)$ for the inference to be valid, aside from truth preservation at all worlds in all models, it is additionally required that the conclusion context index *z* preserves the mutual contextual information of context indices *x* and *y*, which make the premises true. Clearly, this condition is met trivially with Berto's restriction in place. Naturally, the model theory ensures that all the relevant terms such as *context*, *context index*, *contextual information preservation*, and *mutual contextual information* are properly and carefully motivated and defined. So although standard inferences fail to hold in general, when we lift Berto's restriction, all of their instances that are said to *preserve contextual information* (have a suitably contextualized form) do hold.

---

[64] My emphasis. By *Adjunction* Berto means the semantic condition that guarantees the validity of the inference pattern $[A]B, [A]C \vDash [A](B \wedge C)$. I shall refer to it as *Adjunction of Consequents.*

[65] That is, valid on all weakly centered sphere systems, i.e. those characterizing Lewis' logic VW.

1.4.6 Nolan's context-relativization suggestion

A suggestion along similar lines to Berto's can be found in Nolan (1997). It proposes a modification of Stalnaker-Lewis similarity semantics for counterfactuals in a manner that emphasizes the role of context in order to account for the interpretation of *similarity* as *similarity in relevant respects*. Whereas Berto offers advice on addressing the matter syntactically, by accounting for contextual differences explicitly in the object language with an indication of how an index set should feature in a model, Nolan hints at a corresponding modification of similarity assignments that would accommodate such contextual disambiguation. The two suggestions are complimentary (in my view) – jointly amounting to a rudimentary recipe for an analysis of counterfactuals that explicitly accounts for the influence of context.

> A more sophisticated approach would be to employ, instead of a function from worlds to sets of spheres, a function from worlds and contexts to sets of spheres. This may deal better with representing our use of relevant similarity *in determining the spheres*, since what is more relevantly similar than what is often (always?) a matter of context. Contexts themselves are not monolithic, of course, and there is a potential to develop a quite sophisticated formal mechanism for modelling the selection of sets of spheres. (Nolan 1997, n.28)

This will become clearer when I introduce Lewis' similarity sphere models in the next chapter. It will suffice to say at this point that Berto (2017) and Nolan's (1997) suggestions have one thing in common – how to include objects interpreted as contexts into a model theory. The account I give in chapter 4, develops and implements these nascent ideas.

Chapter 2

Conditional logics and David Lewis' analysis of counterfactuals.

> *'If kangaroos had no tails, they would topple over' is true (or false, as the case may be) at our world, quite without regard to those possible worlds where kangaroos walk around on crutches, and stay upright that way. Those worlds are too far away from ours. What is meant by the counterfactual is that, things being pretty much as they are – the scarcity of crutches for kangaroos being pretty much as it actually is, and so on – if kangaroos had no tails they would topple over.*
>
> Lewis (1973, pp.8-9)

2.0 Introduction

In this chapter I give a detailed and critical exposition of Lewis' *similarity semantics* for counterfactuals in terms of his similarity spheres. In the first part of the chapter I give a brief overview of a popular semantics for *ceteris paribus* conditionals. Then I give an account of Lewis' (1973) critique of the strict conditional and his argument that counterfactuals don't correspond to any strict conditional, but rather a variably strict conditional. In the second, and main part of the chapter, I present Lewis' semantics for the counterfactual in terms of similarity sphere systems $ and give an in-depth survey of various conceptions of comparative similarity of worlds and the corresponding restrictions on $. This exposition includes a critical comparison of Lewis' semantics with another well-known *similarity semantics* for conditionals due to Stalnaker (1968, 1970). In the latter part of the chapter I offer a critical reply to Lewis' logical arguments in favor of the vacuous account of counterfactuals with impossible antecedents, i.e. counterpossibles, and show them to be either inconclusive or unconvincing. This reply to Lewis comprises one of three replies to Lewis' defense of vacuous analysis of counterpossibles. The remaining replies are given in *Chapter 3*. Throughout the entire chapter I endorse the system characterized by weakly centered systems of similarity spheres, which is the weakest system that validates *Modus Ponens*, yet strong enough to invalidate problematic inferences burdening Lewis' preferred logic, which is characterized by a system satisfying a stronger centering condition.

2.1 Conditional logic

From a Kripke semantics perspective, conditional logics are modal logics, with a multiplicity of accessibility relations of a certain kind.[66] That is, the semantics for conditional logics extends the conceptual mechanism developed by Kripke. This approach has been motivated by giving a better account of the conditional – both the indicative and the subjunctive – by developing a semantics that invalidates a number of questionable inference forms, such as *transitivity*, *antecedent strengthening*, and *contraposition*, which remain valid in classical logic, for both the material and strict conditionals. Below are the formal patterns of those inferences, accompanied by examples from natural language where they appear to fail.

Transitivity: $A \supset B, B \supset C \vDash A \supset C$

If Hoover had been born in Russia, he would have been a communist.

If Hoover had been a communist, he would have been a traitor.

∴ Hence, if Hoover had been born in Russia, he would have been a traitor.

Antecedent Strengthening: $A \supset B \vDash (A \wedge C) \supset B$

If I strike a match, the match will light up.

∴ Therefore, if I submerge myself in the pool and strike a match, the match will light up.

Contraposition: $A \supset B \vDash {\sim}B \supset {\sim}A$

If I have any coffee, I only have a little.

∴ So, if I have a lot of coffee, I don't have any.

2.1.1 The formal language

Let's start by describing the formal language, relevant to this chapter, and the corresponding set of well-formed formulae over that language. First let us start with the basic ingredients for our language, i.e. a set of propositional variables $PV = \{p_n : n \in \mathbb{N}\}$ the elements of which shall be denoted with lowercase Roman letters $(p, q, r, \dots)$ or subscripted lowercase Roman p's $(p_1, p_2, \dots, p_k, \dots)$, or lowercase Greek letters $(\varphi, \psi, \chi, \dots)$; unary connectives: ∼ (negation), □ (necessity), ◊ (possibility); and binary connectives: ∧ (conjunction), ∨ (disjunction), ⊃ (material conditional). ≺ (strict conditional), > (counterfactual conditional). For the

[66] Priest (2008, p.82). Or rather, treating them as modal logics with such properties of the accessibility relation is one of the ways conditionals can be analyzed. Much of the presentation and discussion of conditional logics, *ceteris paribus* conditionals in particular in this chapter, follows Priest (2008, §5).

metalanguage, upper case letters $(A, B, C, \ldots)$ shall be used as variables ranging over formulae and propositional variables. The recursive formation rules for the set of well-formed formulae (For) of the formal language are given below.

Definition 2.1: Let For be the smallest set closed under the following well-formed formula formation rules:

B: All propositional variables are wffs, i.e. $PV \subseteq For$.

R1: If $A \in For$ then $\{\sim A, \Box A, \Diamond A\} \subseteq For$.

R2: If $\{A, B\} \subseteq For$ then $\{A \wedge B, A \vee B, A \supset B, A \prec B, A > B\} \subseteq For$.

2.1.2 Strict conditionals

It's easy to check that *transitivity*, *antecedent strengthening*, and *contraposition* inference forms are valid for the material conditional. They're also valid for the strict conditional on the class of Kripke frames, as I'll shortly demonstrate. The strict conditional $\prec$, is defined in the following way $A \prec B := \Box(A \supset B)$. The key thing to note here is the fact that the accessibility relation *R* remains invariant with respect to the antecedent on Kripke semantics for the strict conditional. Kripke frames and models provide a point of reference to the sphere frames that Lewis uses in setting up his argument that the counterfactual is not any strict conditional – an argument to the layout of which section §2.2 is devoted. For Kripke semantics for normal modal logics – which I will be referring to often throughout the following few sections – see definitions 1.0 through 1.3, in §1.3.1, the weakest of which is the system K.[67]

Extensions of the basic logic K are achieved by placing various constraints on the accessibility relation R, e.g. reflexivity, symmetry, or transitivity. Each such constraint defines a class of models (a subclass of all Kripke models) characterizing a different conception of necessity corresponding to some normal modal logic. The key thing to note here is that if an inference is valid on all Kripke models, then *a fortiori* it is valid on a subset of Kripke models. In particular, by demonstrating that inference forms such as transitivity, antecedent strengthening, and contraposition are K-valid for the strict conditional, then they are valid for all extensions of K. It should be noted that because on a Kripke model (W, R, V)

[67] I shall not focus here on the constraints on *R* that may be put in place, and instead focus on the relevant distinction here, between a single accessibility relation on Kripke frames and entire families of accessibility relations (indexed by formulae) on frames for conditional logics.

the truth of the strict conditional $A \prec B$ at a world $w \in W$ reduces to the truth of its material counterpart $A \supset B$ at *all* worlds accessible from w, and since the above inferences are valid for the material conditional, the validity for the strict conditional follows. I'll provide a proof of the validity of antecedent strengthening, by way of illustration. The remining inference forms can also be easily shown to be K-valid.

Proposition 2.1: $A \prec B \vDash_K (A \wedge C) \prec B$

Proof : Suppose that $w \Vdash A \prec B$ on some Kripke model (W, R, V) and $w \in W$. So, $w \Vdash \Box(A \supset B)$ by definition of $\prec$. Hence, $\forall u \in W$, such that wRu, $u \Vdash A \supset B$, So, $\forall u \in W$, such that wRu, either $u \Vdash \sim A$ or $u \Vdash B$. If $u \Vdash \sim A$, then not $u \Vdash A$, and therefore not $u \Vdash A \wedge C$, for any $C \in For$. Hence $u \Vdash \sim(A \wedge C)$ for any $C \in For$. So, $u \Vdash (A \wedge C) \supset B$ for any $C \in For$. Now, if $u \Vdash B$, then $u \Vdash D \supset B$ for any $D \in For$. In particular $u \Vdash (A \wedge C) \supset B$. In conclusion, it follows that $\forall u \in W$, such that wRu, $u \Vdash (A \wedge C) \supset B$, as required. □

2.1.3 *Ceteris paribus* conditionals

A way of getting around the validation of those inference forms is to fashion conditional semantics whereby the single accessibility relation R is replaced with a whole family of accessibility relations $\{R_A : A \in For\}$, indexed by formulae. The philosophical motivation here is that not all worlds count as relevant in the evaluation of the conditional at some world, but rather only those worlds that are *ceteris paribus* the same as the world of evaluation. The model-theoretic means of capturing this intuition has been to have R_A access only those worlds where A is true and which are *ceteris paribus* the same as the actual world. So, it's clear that unlike on Kripke frames, for any world, conditionals with different antecedents need not be evaluated on the same accessible worlds. The truth conditions for the *ceteris paribus* conditional $>$ are almost the same as for the strict conditional, save for the newly introduced variability of the accessibility relation: $A > B$ is true at a world w iff at all worlds in $\{u : wR_A u\}$ $A \supset B$ is true. And since $\{u : wR_A u\} = \{u : wR_B u\}$ need not be true in general, for any world w and distinct antecedents A and B, the problematic inference forms are invalidated.[68] Before I continue discussing other key differences of this approach to Kripke semantics, I'll present the formal definition of the frames and models underlying the semantics of thusly conceived *ceteris paribus* conditionals.

[68] See (Priest 2008, §5.2-5.3).

Definition 2.2: A *C frame* is a pair $(W, \{R_A: A \in For\})$ where W is a nonempty set, and for each $A \in For$, $R_A \subseteq W \times W$.[69] Formally, W is an arbitrary set of objects. On the intended interpretation, relevant to the semantics under consideration, its elements are as *possible worlds*. R_A is still called the accessibility relation, just like on Kripke frames, with one obvious key addition regarding the interpretation of the formula index. Intuitively wR_Au means that u is an A-world accessible from w, which is *ceteris paribus*, the same as w.

Definition 2.2.1: For convenience, define $f_A(w) := \{u: wR_Au\}$, i.e. the set of worlds accessible from w under R_A.

Definition 2.2.2: It will also be convenient to define $[A]^{\mathcal{M}} := \{w \in W: \mathcal{M}, w \Vdash A\}$ for any model $\mathcal{M}$ with domain W, and any class of models discussed in this chapter. The superscript will be often omitted in cases when its absence will not lead to ambiguity.

Definition 2.3: A *C model* is a triple $(W, \{R_A: A \in For\}, V)$, where $(W, \{R_A: A \in For\})$ is a *C frame*, as defined earlier, and $V: PV \longrightarrow \wp(W)$, is the function that assigns to each propositional variable $p \in PV$ a subset of W. Informally we think of $V(p)$ as the set of worlds in the model where is p true.
Truth in a model is defined in terms the satisfiability relation $\Vdash \subseteq W \times For$. We read $w \Vdash A$ as 'A is true at w'. Given a C-model $(W, \{R_A: A \in For\}, V)$ and any $w \in W$, define $\Vdash$ as follows:

(1) – (6) are as for Kripke models.

(7) $w \Vdash A > B$ iff $\forall u \in W$, such that wR_Au, $u \Vdash B$.

Equivalently we can express the truth conditions for $A > B$ more concisely in terms of previously defined sets of worlds:

(7') $w \Vdash A > B$ iff $f_A(w) \subseteq [B]$.

That is, iff B is true at all R_A-accessible worlds.

[69] C stands for *conditional*. I'm basing the frame theory and model theory for conditional logics on Priest (2008, §5.3). What I call *C frames* are special cases of structures widely known among computer scientists as *labelled transition systems* $(W, \{R_a: a \in A\})$, where $W \neq \emptyset$ is a set of states, and $A \neq \emptyset$ is a set of labels (Blackburn 2001, p.3).

Definition 2.4: C validity

Let $\vDash \subseteq \wp(For) \times For$, and define $\Sigma \vDash_C A$ iff for all Kripke models $(W, \{R_A: A \in For\}, V)$, and all $w \in W$, if $w \Vdash B$ for all $B \in \Sigma$, then $w \Vdash A$. That is, valid inference is defined as truth preservation at all worlds in all *C models*. A formula $A \in For$ is said to be valid iff it is true in all C models (notation: $\emptyset \vDash_C A$).

Priest (2008, §5.3) calls the logic characterized by the class of *C models* conditional logic C. Since no constrains are placed on the relations R_A, C is the analogue for conditional logics of the modal logic K. Below is a counterexample to antecedent strengthening in C.

Proposition 2.2: $p > q \nvDash_C (p \wedge r) > q$

Proof : Consider the following countermodel: $W = \{w, u\}, R_p = \emptyset, R_{p \wedge r} = \{(w, u)\}$; for all other $A \in For$, R_A can be anything; and V is such that $u \notin [q]$. Now $w \Vdash p > q$, since $\emptyset = f_p(w) \subseteq [A]$ for any $A \in For$, but $w \nVdash (p \wedge r) > q$, since $\{u\} = f_{p \wedge r}(w) \nsubseteq [p]$.[70] □

However, C is a rather weak logic, as it doesn't even validate *Modus Ponens* for >. There's an important extension of C, which is strong enough to validate *Modus Ponens* for >, while being sufficiently weak to invalidate the aforementioned questionable inference forms.[71] An important extension of C is the logic C+, which results from placing additional constraints on R_A.

Definition 2.5: A *C+ model* is a C model where for each $A \in For$ the accessibility relation R_A satisfies the following additional constraints:

(1) $f_A(w) \subseteq [A]$
All worlds that are R_A-accessible from w, are A-worlds.

(2) If $w \in [A]$ then $w \in f_A(w)$
If A holds at a world then that world is self R_A-accessible.

One may impose additional conditions on R_A thereby generating other extensions of C. This is what we turn to now – discussing the general approach suggested by Stalnaker and Lewis

[70] I have borrowed this counter-model from Priest (2008, pp.86-87).
[71] Giving the corresponding proofs is beyond the relevant scope of this chapter. For a comprehensive discussion of C and C+ and their respective proof theories, see Priest (2008, §5.4-5.5).

in their seminal work on conditional logics. The next section gives a detailed account of Lewis' similarity spheres semantics, which begins with an outline of Lewis' reasoning and motivations that led him to the formulation of that approach to analyzing the counterfactual. Most of that is captured in my critical paraphrase of Lewis' argument why the counterfactual can't be any strict conditional. The manner in which Lewis chooses to express that argument serves also as an excellent introduction to some of the formal foundations of his systems of similarity spheres semantics.[72]

2.2 Lewis' general proposal for counterfactuals

2.2.1 Why the counterfactual is not a strict conditional

Lewis (1973) suggests a possible-world semantics for the counterfactual conditionals based on the idea of *variably strict necessity*, conceived as variable *overall similarity of worlds*. That account, being a milestone in the work on counterfactuals, and owing to its remarkably intuitive appeal has been a popular starting point to recent directions of research containing proposals for *non-vacuous* treatments of counterpossibles, unlike on Lewis' own account (more on that later). I will presently cover Lewis' motivations for claiming that the counterfactual is more like a variably strict conditional, since – as he argues – it can't be any single strict conditional.

Lewis argues that since in general the necessity operator acts like a restricted universal quantifier over possible worlds, necessity of a certain kind is just truth at worlds that satisfy some restriction. Such worlds are called *accessible*, in the sense of satisfying the conditions of the necessity under consideration.[73] For example, *physical necessity* is truth at worlds satisfying the accessibility restriction of having the actual physical laws hold at them.
For the purposes of his argument that *the counterfactual is not any single strict conditional* Lewis defines an alternative, yet clearly equivalent (which I'll prove) class of frames to Kripke's – a move that aims to shift emphasis from talking about various *accessibility relations* to talking about various *spheres of accessibility* (essentially, such spheres are just images of worlds under the accessibility relation) when modelling necessity. That picture aims to make salient the role of necessity operators as restricted universal quantifiers over possible worlds, thereby facilitating talk of *variable necessity*, which is the key idea

[72] For some interesting, current work on *ceteris paribus* conditionals see Girard & Triplett (2018).
[73] The reminder of this section is my paraphrase of Lewis' discussion given in (Lewis 1973, pp.5-9).

underlying the notion of *strict conditionals of varying strictness*, and a foundational step to developing the notion of a *variably strict conditional*, as a model for the *counterfactual*. The next few definitions are devoted to that class of models. Aside from aiding his argument in the manner described above, these models are also the first conceptual step toward formulating Lewis' sphere semantics for counterfactuals, which I precisely define and discuss in the next section.[74] What follows is a detailed paraphrase of Lewis' argument.

Definition 2.6: A *sphere frame* is a pair (W, S), where W is as for Kripke frames, and $S: W \longrightarrow \wp(W)$.[75] For each $w \in W$, the set S_w is called the *sphere of accessibility* around w, regarded as the set of worlds accessible from w. So, '$u \in S_w$' is read as 'u is accessible from w'.

It should be noted that Sphere frames are essentially equivalent to Kripke frames. Roughly, whereas on Kripke frames R is an arbitrary binary relation on W, on sphere frames, each S_w corresponds to the image of w under R, and S is just the set $\{(w, S_w): w \in W\}$. A precise argument for this correspondence is given below, i.e. *Lemma 2.3*.

Definition 2.7: A *sphere model* is the triple (W, S, V), where (W, S) is a sphere frame, and V is as for Kripke models.
Truth in a model is defined in terms the satisfiability relation $\Vdash \subseteq W \times For$. We read $w \Vdash A$ as 'A is true at w'. Given a sphere model (W, S, V) and any $w \in W$, define $\Vdash$ as follows:

(1) – (5) are as for Kripke models in *Definition 1.2*.

(6) $w \Vdash \Box A$ iff $\forall u \in S_w: u \Vdash A$.

Lemma 2.3: The exists a one-to-one correspondence $h: \mathbf{K} \longrightarrow \mathbf{S}$ between the classes of Kripke frames **K** and sphere frames **S**, such that $\mathfrak{F}$ is isomorphic to $h(\mathfrak{F}) \in \mathbf{S}$ for each frame $\mathfrak{F} \in \mathbf{K}$.

Proof : First to prove (1). I'll proceed by showing that there exist injections: $f: \mathbf{K} \longrightarrow \mathbf{S}$ and $g: \mathbf{S} \longrightarrow \mathbf{K}$, between **K** as and **S**, such that if $f(\mathfrak{F}_\mathbf{K}) = \mathfrak{F}_\mathbf{S}$, then $g(\mathfrak{F}_\mathbf{S}) = \mathfrak{F}_\mathbf{K}$, and if $g(\mathfrak{F}_\mathbf{S}) = \mathfrak{F}_\mathbf{K}$, then $f(\mathfrak{F}_\mathbf{K}) = \mathfrak{F}_\mathbf{S}$, for any $\mathfrak{F}_\mathbf{K}$ and $\mathfrak{F}_\mathbf{S}$. That is, $g\big(f(\mathfrak{F}_\mathbf{K})\big) = f\big(g(\mathfrak{F}_\mathbf{S})\big)$, for any $\mathfrak{F}_\mathbf{K}$ and $\mathfrak{F}_\mathbf{S}$. This will justify the existence of the bijection $h: \mathbf{K} \longrightarrow \mathbf{S}$ defined as $h = f$ and $h^{-1} = g$.

[74] The entire discussion and setting up of the argument which I'm formally rephrasing here, plus the argument it leads up to, is to be found in (Lewis 1973, pp.5-9).
[75] These are not to be confused with sphere systems, discussed in the next section. Lewis (1973, p.7) states that the sphere formulation is *obviously* equivalent to the general (Kripke) semantics, but I demonstrate the required correspondence.

To compete the proof, it will also be shown that h satisfies the *isomorphic property*, that is, $h(R|_w) = S_{h(w)}$, by showing that $f(R|_w) = S_{f(w)}$ and $g(S_w) = R|_{g(w)}$, for all $w \in W$.

Definition: for a relation $R \subseteq W^2$, let the *image of w under R* be the set $R|_w := \{u: wRu\}$.

Definition: Let the map $f: \mathbf{K} \longrightarrow \mathbf{S}$ be defined as follows:

(i) $f(W) = W$.

(ii) $f(R|_w) = S_w = \{u: wRu\}$, and $f(R) = \{(w, S_w): w \in W\}$.

Definition: Let the map $g: \mathbf{S} \longrightarrow \mathbf{K}$ be defined as follows:

(i) $g(W) = W$.

(ii) $g(S_w) = R|_w = \{(w, u): u \in S_w\}$, and $g(S) = \cup\{R|_w: w \in W\}$.

Definition: Let the map $h: \mathbf{K} \longrightarrow \mathbf{S}$, be as follows: $h = f$ and $h^{-1} = g$.

Now to show that if $f(R) = S$, then $g(S) = R$, and if $g(S) = R$, then $f(R) = S$, for any R and S, that is $f \circ g = g \circ f = id$, for any R and S, which will justify the definition of the *bijection h* as $h = f$ and $h^{-1} = g$. The functions are clearly injective.

Start with some Kripke frame (W, R) with $R \subseteq W \times W$, and let $f(R) = \{(w, S_w): w \in W\}$, such that for each $w \in W$, $f(R|_w) = S_w = \{u: wRu\}$. Now we consider the sphere frame $\big(f(W), f(R)\big)$ obtained this way, where $S = f(R): f(W) \longrightarrow \wp\big(f(W)\big)$, and show that $g\big(f(R)\big) = R$. We start by establishing that $g\big(f(R|_w)\big) = g(S_w) = R|_w$ for each $w \in W$. Clearly $R = \cup\{R|_w: w \in W\}$, but since $S = f(R)$, it follows from the definition of g that $g\big(f(R)\big) = g(S) = \cup\{R|_w: w \in W\} = R$. □

Now, we start with some sphere frame (W, S) with $S: W \longrightarrow \wp(W)$, and let $g(S) = \cup\{R|_w: w \in W\}$, such that $g(S_w) = R|_w = \{(w, u): u \in S_w\}$ for each $w \in W$. Now we consider the Kripke frame $\big(g(W), g(S)\big)$ obtained this way, where $R = g(S) \subseteq g(W) \times g(W)$, and show that $f\big(g(S)\big) = S$. We start with noting that $f\big(g(S_w)\big) = f(R|_w) = S_w$ for each $w \in W$. Clearly $S = \{(w, S_w): w \in W\}$, but since $R = g(S)$, it follows that $f\big(g(S)\big) = f(R) = \{(w, S_w): w \in W\} = S$, by definition of f. □

Theorem 2.3.1: To any Kripke model $\mathcal{M}$ there corresponds a unique sphere model $\mathcal{M}'$ such that for all $w \in W$ and $A \in For$: $\mathcal{M}, w \Vdash A$ iff $\mathcal{M}', w \Vdash A$. Extend h in the following way:

(iii) $f(V) = g(V) = V$

Now h has been extended to a bijection between Kripke models and sphere models. The corresponding models have the same domain and same value assignments to propositional variables. Only R and S differ. Note that the only difference between Kripke models and sphere models are the truth conditions for $\Box$:

For *Kripke models*: $w \Vdash \Box A$ iff $\forall u \in W$, such that wRu, $u \Vdash A$.
For *sphere models*: $w \Vdash \Box A$ iff $\forall u \in S_w$: $u \Vdash A$.

Now assume $\mathcal{M}, w \Vdash \Box A$ for some Kripke model $\mathcal{M} = (W, R, V)$. Then by definition $\forall u \in W$, such that wRu, $u \Vdash A$. Now, consider $\mathcal{M}' = f(\mathcal{M})$. Now, it will be shown that $f(\mathcal{M}), w \Vdash \Box A$, on the assumption that $\forall u \in W$, such that wRu, $u \Vdash A$. By definition, $f(R|_w) = \{u: wRu\}$, which consists of all worlds $u \in W$, such that wRu. But if for all those worlds $u \Vdash A$, by hypothesis, then $u \Vdash A$ for all $u \in f(R|_w)$. Hence $f(\mathcal{M}), w \Vdash \Box A$, as required. □

Now assume $\mathcal{M}', w \Vdash \Box A$ on the sphere model $\mathcal{M}' = f(\mathcal{M}) = \big(f(W), f(R), f(V)\big)$. We need to show that $g\big(f(\mathcal{M})\big), w \Vdash \Box A$. Note that, since $f(\mathcal{M}) = \mathcal{M}'$, then $g\big(f(\mathcal{M})\big) = g(\mathcal{M}')$. So, now we start by assuming $\mathcal{M}', w \Vdash \Box A$ and need to show that $g(\mathcal{M}'), w \Vdash \Box A$. Assuming $\mathcal{M}', w \Vdash \Box A$, if follows that $u \Vdash A$ for all $u \in S_w$. Now, $g(S_w) = R|_w = \{(w, u): u \in S_w\}$, so given the hypothesis, it follows that for all $(w, u) \in R$, if $(w, u) \in g(S_w)$ then $u \Vdash A$. That is, if wRu then $u \Vdash A$. Therefore, $g(\mathcal{M}'), w \Vdash \Box A$, as required. This completes the proof. □

2.2.1.1 Various kinds of necessity

Lewis' argument hinges on the demonstration that the counterfactual can't be any single strict conditional. But in order to define strict conditionals of *varying strictness*, necessity operators corresponding to varying restrictions of the universal quantifier over possible worlds need to be defined. Variable necessity operators can be modelled in terms of correspondingly varying spheres of accessibility as the intended restrictions of the universal quantifier over worlds. There are infinitely many kinds of restrictions that one may impose on the range of accessibility for any world *w*, and to each such restriction there corresponds a necessity operator. I'll gradually make the notion of such restrictions (and variable necessity operators) more precise. Let's start with an intuitive description.

Example

Here are a few *examples* of such restrictions, relevant to the current discussion:

- Logical necessity: S_w is restricted to logically possible worlds.

- Physical necessity: S_w is restricted to worlds where the laws of nature prevailing at w hold.
- Necessity in respect of facts of so-and-so-kind: for example, S_w is restricted to worlds where kangaroos have no tails.

To each restriction there corresponds a necessity operator: $\Box^L$ corresponding to logical necessity, $\Box^P$ corresponding to physical necessity, $\Box^K$ corresponding to necessity in respect of facts of so-and-so-kind (e.g. kangaroos not having tails).

Definition 2.8: Denote the *index set* of all restrictions on the range of accessibility with $\mathcal{I}$. Intuitively, each $i \in \mathcal{I}$ corresponds to some kind of necessity, e.g. logical, physical, etc.

This gives rise to a class of operators $\{\Box^i : i \in \mathcal{I}\}$. Intuitively, each $\Box^i$ is a necessity operator corresponding to some kind of necessity $i \in \mathcal{I}$. Each such operator corresponds to a restriction of the universal quantifier over possible worlds to i-possible worlds. The next few definitions make this intuitive description more precise.
To accommodate the new class of connectives, we need to expand the formal language described in §2.1.1 by the addition of the set of operators $\{\Box^i : i \in \mathcal{I}\}$, and the class of abbreviations for the corresponding strict conditionals $\{>^i : i \in \mathcal{I}\}$, thereby extending the set For of well-formed formulae in the following way:

R3: If $A \in For$ and $i \in \mathcal{I}$, then $\Box^i A \in For$.
R4: If $\{A, B\} \subseteq For$ and $i \in \mathcal{I}$, then $A >^i B \in For$.

In order to accommodate the expansion of our language, the original sphere frames need to be modified accordingly, by admitting sphere functions that correspond to each kind of variable necessity operator.

Definition 2.9: A *variable sphere frame* is a pair $(W, \{S^i : i \in \mathcal{I}\})$ where W is a nonempty set of objects, and $\mathcal{I}$ is a nonempty set of indices, and $S^i : W \longrightarrow \wp(W)$ for each $i \in \mathcal{I}$. We can denote $\{S^i : i \in \mathcal{I}\}$ with $\mathcal{S}$, for brevity when its content is unambiguous.[76]

On the intended interpretation, the elements of W are possible worlds, $\mathcal{I}$ is as given in *Definition 2.8*, and for each $w \in W$, the set S^i_w is called the i-sphere of accessibility around w,

[76] Thus variable sphere frames are sets paired with families of their subsets.

which is regarded as the set of worlds accessible from *w* corresponding to the restriction of the universal quantifier to *i*-possible worlds, relative to *w*.

Definition 2.10: A *variable sphere model* is a triple $(W, \mathcal{S}, V)$, where $(W, \mathcal{S})$ is a variable sphere frame, and V is as for Kripke models.
Truth in a model is defined in terms the satisfiability relation $\Vdash \subseteq W \times For$. We read $w \Vdash A$ as 'A is true at w'. Given a variable sphere model $(W, \mathcal{S}, V)$ and any $w \in W$, define $\Vdash$ as:

(1) – (5) are as for Kripke models in *Definition 1.2*, (the non-modal part of the language)
(6) $w \Vdash \Box^i A$ iff $\forall u \in S_w^i$: $u \Vdash A$.

2.2.1.2 Strict conditionals of varying strictness

Now we're ready to define a whole class of strict conditionals of varying strictness. Given a variable sphere model $(W, \mathcal{S}, V)$ and any $w \in W$, define $\Vdash$ as follows:
(7) $w \Vdash \Box^i(A \supset B)$ iff $\forall u \in S_w^i$: $u \Vdash A \supset B$.

Notation: Denote $\Box^i(A \supset B)$ with $A >^i B$ for each $A, B \in For$ and $i \in \mathcal{I}$.

Definition 2.11: Hierarchy of strictness. For all $i, j \in \mathcal{I}$ and all $A, B \in For$:
The conditional $A >^i B$ is *stricter at world w* than $A >^j B$ iff $S_w^j \subsetneq S_w^i$. The conditional $A >^i B$ is *stricter* than $A >^j B$ iff for all $w \in W$, $A >^i B$ is *stricter at world w* than $A >^j B$.

2.2.1.3 The intended model

The discussion throughout next section, mainly in the formulation of Lewis' argument that no strict conditional is adequate to model the counterfactual will be done with reference to the *intended* variable sphere model $\mathcal{M}_0 = (W^{\mathcal{M}_0}, \mathcal{S}^{\mathcal{M}_0}, V^{\mathcal{M}_0})$, where $W^{\mathcal{M}_0}$ is the set of *all* possible worlds. Among its accessibility assignments, on top of those corresponding to logical necessity S^L, $\mathcal{S}^{\mathcal{M}_0}$ also contains those corresponding to other kinds of necessity, like physical necessity S^P, and a whole plethora of necessities in respect of facts of so-and-so-kind, e.g. kangaroos not having tails S^K. On the intended model, $S_w^L = W^{\mathcal{M}_0}$ for all $w \in W^{\mathcal{M}_0}$, whereas S_w^P is the set of possible worlds where the laws of nature (physical laws) prevailing at *w* hold. Finally, S_w^K is the set of possible worlds where the sentence K: 'kangaroos have no tails' is true i.e. $S_w^K = [K]^{\mathcal{M}_0}$. As far as necessity in respect of facts of so-and-so-kind, this generalizes to $S_w^A = [A]^{\mathcal{M}_0}$ for any facts expressed by $A \in For$.

Truth conditions for various strict conditionals on the *intended* model are as on any variable sphere models. Now those various kinds of necessity that only received an intuitive description in the example at the beginning of §2.2.1.1, can now be defined precisely:

(1) – (5) are as for Kripke models in *Definition 1.2*, (the non-modal part of the language)

(6) $\mathcal{M}_0, w \Vdash \Box^i A$ iff $\forall u \in S_w^i$: $\mathcal{M}_0, u \Vdash A$.

(7) $\mathcal{M}_0, w \Vdash \Box^i (A \supset B)$ iff $\forall u \in S_w^i$: $\mathcal{M}_0, u \Vdash A \supset B$.

Some special kinds of various necessity operators:

$\mathcal{M}_0, w \Vdash \Box^L A$ iff $\forall u \in S_w^L$: $\mathcal{M}_0, u \Vdash A$.

$\mathcal{M}_0, w \Vdash \Box^P A$ iff $\forall u \in S_w^P$: $\mathcal{M}_0, u \Vdash A$.

$\mathcal{M}_0, w \Vdash \Box^K A$ iff $\forall u \in S_w^K$: $\mathcal{M}_0, u \Vdash A$.

To tie the above with the notion of varying strictness, note that, $A >^P B$ is *stricter at the actual world* than $A >^K B$ on the intended model, and $A >^L B$ is *stricter* than $A >^K B$ on the intended model.

2.2.1.4 Strict conditionals and comparative similarity of worlds

The counterfactual proposed by Lewis, is closely related to *ceteris paribus* conditionality, and as such is not fundamentally different from the logics C and C+ discussed earlier. The difference lies in how the *ceteris paribus* clause is explained. It is evident from the passage below, that the kind of strict conditional Lewis has in mind is effectively based on accessibility that satisfies certain *ceteris paribus* constrains, which he then suggests are best expressed in terms of *comparative similarity* of worlds.

> Counterfactuals are related to a kind of strict conditional based on comparative similarity of possible worlds. A counterfactual $\varphi > \psi$ is true at a world *w* if and only if ψ holds at certain φ-worlds; but certainly not all φ-worlds matter. '*If kangaroos had no tails, they would topple over*' is true (or false, as the case may be) at our world, quite without regard to those possible worlds where kangaroos walk around on crutches, and stay upright that way. Those worlds are too far away from ours. What is meant by the counterfactual is that, things being pretty much as they are –the scarcity of crutches for kangaroos being pretty much as it actually is, and so on – if kangaroos had no tails they would topple over. Lewis (1973, pp.8-9)

Lewis also observes that in our consideration regarding what kind of restricted necessity should underlie the strict conditional that best captures counterfactual reasoning, aside from ruling out worlds that are grossly dissimilar from the actual world (e.g. where tailless kangaroos use crutches to stay upright), we also must avoid deeming as accessible worlds that are *too similar* (or at least be careful when doing so). For if we include into the accessibility sphere worlds where kangaroos have no tails, but otherwise *everything else is exactly the same as the actual world*, then kangaroos despite being tailless would nevertheless leave tail tracks in the sand, and their genetic make-up, being as it actually is, would nevertheless somehow still code for different phenotypical traits (absence of tail). Lewis concludes that counterfactuals are apparently based on a strict conditional corresponding to an accessibility assignment determined by an *overall similarity of worlds*, where respects of difference and respects of similarity are "somehow" balanced off against each other.

In the light of this, let's consider again our earlier example of tailless-kangaroos necessity. It is clear from the above concerns regarding the intended notion of comparative similarity of worlds that the strict conditional $\Box^K(K \supset T)$ would fall short of being the adequate model for the evaluation of the counterfactual 'If ***K****angaroos had no tails*, they would ***T****opple over*'. This inadequacy stems from the fact that the restriction corresponding to $\Box^K$, namely $[K]^{\mathcal{M}_0}$, includes *all* possible K-worlds, and as such it may include worlds that have just been argued to be "too far" to be regarded as relevant to the evaluation of the counterfactual in question.

What we require, is a restriction that meets the similarity criteria that Lewis has in mind. That is, we are interested in $s \in \mathcal{I}$ corresponding to an accessibility restriction $S_w^s \in \mathcal{S}^{\mathcal{M}_0}$ determined by an *overall similarity of worlds* such that $S_w^s \cap [K]^{\mathcal{M}_0}$ doesn't contain worlds where kangaroos manage to stay upright with the aid of crutches, or worlds where kangaroos leave a tail track behind in the sand as they actually do (despite their taillessness). That is, all K-worlds in S_w^s are sufficiently similar to w to count as relevant in the evaluation of the counterfactual with antecedent K at w. Then the strict conditional $\Box^s(K \supset T)$ would meet the accessibility assignment requirement as determined by an *overall similarity of worlds* in the manner intended by Lewis. It's apparent that Lewis intends there to be a fitting restriction $s \in \mathcal{I}$ of *this kind* in general, for any antecedent $A \in For$ of any strict conditional based on comparative similarity of worlds.

This would yield the following account of the counterfactual when considered as a single strict conditional based on the comparative similarity of worlds, i.e. the intended model truth conditions for the counterfactual in terms of a single strict conditional based on the comparative similarity of worlds are as follows:

Definition 2.12: The counterfactual 'If A, then B' is true at a world w iff $\mathcal{M}_0, w \Vdash A >^s B$.

Recalling that $\mathcal{M}_0, w \Vdash A >^s B$ iff $\forall u \in S_w^s$: $\mathcal{M}_0, u \Vdash A \supset B$ for all $A, B \in For$, $w \in W$, and where all worlds in S_w^s are regarded as sufficiently similar to w to count as relevant in the evaluation of the counterfactual with antecedent A at w.

To be precise, the *intended* accessibility restriction S_w^s corresponding to $s \in \mathcal{I}$ isn't constant, but rather a function of the antecedent. Lewis' argument that the counterfactual cannot be any strict conditional amounts to showing that this can't be done on the intended model, where we are only equipped with a class of strict conditionals of fixed strictness. The accessibility restriction index denoted with s, used throughout the discussion in the next section is just to remind us that we're not talking about arbitrary elements of $\mathcal{I}$, but ones that correspond to accessibility restrictions determined by an *overall similarity of worlds*. I devote the next section to Lewis' argument, and given that it is expressed with direct reference to the intended model $\mathcal{M}_0$, I shall omit any model-denoting superscripts.

2.2.1.5 The argument

Lewis gives the following argument in support of the claim that no single strict conditional $A >^s B$, fashioned in the manner just discussed, is adequate to model the counterfactual: counterfactuals, modelled this way – by any single strict conditional – fare well when considered in isolation from other counterfactuals, but problems arise if several counterfactuals are considered together.[77] Naturally, this is unacceptable if an *adequate* analysis is to be given for a language that contains conjunctions. The argument makes use of the fact that the antecedent-strengthening rule, which is valid for strict conditionals – see *Proposition 2.1* – and *a fortiori* for strict conditionals corresponding to an accessibility assignment determined by an overall similarity of worlds. Suppose we're using a counterfactual modelled by any *single* strict conditional '$>^s$' of some fixed strictness (as

[77] Paraphrasing Lewis (1973, pp.910).

defined above, in *Definition 2.12*). Lewis invites us to consider the following list of counterfactuals, with their respective translations into the object language given below:[78]

If I walked on the lawn, no harm would come of it.
$p_1 >^s q$

If I walked on the lawn and everyone did that, the lawn would be ruined.
$(p_1 \wedge p_2) >^s \sim q$

If I walked on the lawn and everyone did that, but everyone was careful, no harm would come of it.
$(p_1 \wedge p_2 \wedge p_3) >^s q$

If …

This sequence of counterfactuals listed below on the left, is accompanied by a corresponding list of their respective negated opposites on the right, to aid the argument.[79] For the kind of cases that are being considered, both the counterfactual and its negated opposite are held true.

(1)	$p_1 >^s q$	and	$\sim(p_1 >^s \sim p)$
(2)	$(p_1 \wedge p_2) >^s \sim \psi$	and	$\sim\big((p_1 \wedge p_2) >^s q\big)$
(3)	$(p_1 \wedge p_2 \wedge p_3) >^s \psi$	and	$\sim\big((p_1 \wedge p_2 \wedge p_3) >^s \sim q\big)$
:	:		

We can imagine prolonging such sequences to arbitrary length. All the above counterfactuals are intuitively true. However, it's evident that they can't *all* be true on any *single* strict conditional of fixed strictness. That is, it should be observed that each counterfactual (on the left) at each stage *n*, contradicts the opposite of the counterfactual at stage *n*+1. That is, the expression in the left column at stage *n* contradicts the expression in the right column at stage *n*+1.

Consider, the first two stages. No matter what degree of similarity of worlds is captured by S_w^s, if $w \Vdash p_1 >^s q$, then q is true in all p_1-worlds in S_w^s. That is, $S_w^s \cap [p_1] \subseteq [q]$, so in particular q is true in all $(p_1 \wedge p_2)$-worlds in S_w^s, i.e. $S_w^s \cap [p_1 \wedge p_2] \subseteq [q]$, since $S_w^s \cap [p_1] \cap [p_2] \subseteq S_w^s \cap [p_1]$ and $S_w^s \cap [p_1 \wedge p_2] = S_w^s \cap [p_1] \cap [p_2]$. So, if the counterfactual is any strict

[78] Lewis (1973, p.10). The negated opposites (listed on the right of each counterfactual) are included since the cases Lewis has in mind are such that they're also hold.
[79] Lewis refers to $A > \sim B$ as the *opposite* of $A > B$, and vice versa, $A > B$ as the *opposite* of $A > \sim B$.

conditional $>^s$, then $p_1 >^s q$ implies $(p_1 \wedge p_2) >^s q$ and contradicts $\sim((p_1 \wedge p_2) >^s q)$. This argument holds for any $s \in \mathcal{I}$, as a corollary of *Proposition 2.1*.

Therefore – Lewis concludes – the counterfactual is not any strict conditional, and he uses this argument as motivation to suggest a *variably strict* conditional as the better alternative for the account of the counterfactual. In the next section I give a detailed account of that proposal.

2.2.2 Counterfactuals as *variably strict* conditionals

In the last section we gave an overview of Lewis' reasons for claiming that strict conditionals best corresponding to counterfactuals are those whose accessibility assignment (conceived of as a *sphere* of accessibility) is determined by an *overall similarity of worlds*, and we also saw Lewis' argument that no *single* strict conditional suffices to give an adequate account of the counterfactual. This means that *a fortiori* no single strict conditional whose accessibility assignment is determined by an *overall similarity of worlds* suffices to give an adequate account of the counterfactual. In this section I present Lewis' solution to this in the form of a *variably strict* conditional account of the counterfactual, and then I'll move onto giving a comprehensive overview (and discussion) of various conceptions of comparative similarity in terms of the model theory offered by Lewis for that account.

The systems of similarity spheres model theory developed by Lewis (1973) is a proposal to analyze he counterfactual as a variably strict conditional, rather than any single strict conditional of fixed strictness. On that novel proposal, instead of a *single sphere* of accessibility, being a subset of W, for each world and $i \in \mathcal{I}$ – as in the case of sphere frames (W, S) or variable sphere frames $(W, \mathcal{S})$ – each world is assigned a *set of accessibility spheres* $\$_w \subseteq \wp(W)$ that satisfy certain restrictions. Each $\$_w$ is a *variable accessibility* assignment. On this picture, the criteria deeming worlds as relevant to the evaluation of a counterfactual at some world *w* is expressed in terms of comparative overall similarity of worlds, whereby any given sphere of accessibility $S \in \$_w$, determined by such considerations, is thought to contain worlds that are similar to *w* to some fixed degree. This way, in conjunction with other spheres of accessibility $S' \in \$_w$ a basis for comparative similarity of worlds to *w* is established. The intuition here is that worlds in some sphere $S \in \$_w$, around some world *w*, are more similar to *w* than those worlds that are outside S. Below is an excerpt from Lewis that captures the heart

of the idea of systems of similarity spheres as representations of comparative similarity of worlds:

> The system of spheres used in interpreting counterfactuals is meant to carry information about the comparative overall similarity of worlds. Any particular sphere around a world *w* is to contain just the worlds that resemble *w* to at least a certain degree. This degree is different for different spheres around *w*. The smaller the sphere the more similar to *w* must a world be to fall within it. [W]henever one world lies within some sphere around *w* and another world lies outside that sphere, the first world is more closely similar to *w* than the second. (Lewis 1973, p.14)

Before giving the formal definition, we can illustrate the core idea of such *systems of spheres* with the aid of the notion of a hierarchy of strictness, introduced earlier in *Definition 2.11*. This illustration will employ elements of Lewis' explanation of the *ceteris paribus* clause in terms of comparative similarity of worlds, and key observations he made in the argument, which has been discussed at the end of the previous section. Consider the two counterfactuals (1) and (2) given below, which I'll presently use to set up a scenario highlighting the link between Lewis' view that counterfactuals are strict conditionals corresponding to an accessibility assignment determined by an overall similarity of worlds, and the limitations of strict conditionals to meet this task (as discussed at the end of the previous section). In this example I'll only rely on the first two stages employed in Lewis' argument, i.e. with the stipulation that all (i), (ii), and (iii) are true (which agrees with intuition).

(1) *If I walked on the lawn, no harm would come of it.*

(i) $p_1 >^s q$

(2) *If I walked on the lawn and everyone did that, the lawn would be ruined.*

(ii) $(p_1 \wedge p_2) >^s \sim q$ and (iii) $\sim\big((p_1 \wedge p_2) >^s q\big)$

If the above counterfactuals were analysed as any strict conditional, then the stipulated scenario would be impossible, since if $p_1 >^s q$ is true, then so is $(p_1 \wedge p_2) >^s q$, by *Proposition 2.1*, which contradicts $\sim\big((p_1 \wedge p_2) >^s q\big)$. In other words, there is no variable sphere model $(W, \mathcal{S}, V)$ such that all (i)-(iii) are true at some world $w \in W$. However, we could assume (1) as true, based on the strict conditional $>^s$ that corresponds to an accessibility assignment S_w^s determined by an overall similarity such that worlds where

'everyone else walks on the lawn', denoted $[p_2]$, would be disregarded as irrelevant to its evaluation at world w. That is, the choice of $s \in \mathcal{I}$ would be such that $S_w^s \cap [p_2] = \emptyset$ for the same reasons as worlds where kangaroos walk upright with the aid of crutches would be disregarded as irrelevant to analyzing the counterfactual *'If kangaroos had no tails, they would topple over'.*

As for (2), we could analyse it by a stricter conditional than $>^s$ (recall *Definition 2.11*), denoted $>^{s'}$, that corresponds to an accessibility assignment $S_w^{s'}$ determined by overall similarity of worlds such that $S_w^s \subseteq S_w^{s'}$ and $S_w^{s'} \cap [p_2] \neq \emptyset$, and all the p_2-worlds in $S_w^{s'}$ are $\sim q$-worlds. On a *system of spheres*, we would analyse the counterfactuals (1) and (2) in the above scenario by a variably strict conditional $>$, fashioned in a manner so that the evaluation of the counterfactual (1) is done in terms of $>^s$ at some world $w \in W$, and the evaluation of the counterfactual (2) at w is done in terms of $>^{s'}$. In other words, a *system of spheres* model $(W, \$, V)$ for the variably strict conditional can be fashioned such that the *variable accessibility* assignment $\$_w = \{S_w^s, S_w^{s'}\}$ satisfies the intended comparative similarity relationship $S_w^s \subseteq S_w^{s'}$, thereby allowing all $p_1 > q$, $(p_1 \wedge p_2) > \sim q$ and $\sim\big((p_1 \wedge p_2) > q\big)$ to be true at $w \in W$. The intuition that worlds in S_w^s are more similar to w than those in $S_w^{s'} \backslash S_w^s$ appears to be preserved, because on the supposition that actually nobody walks on the lawn, it seems strongly intuitive that worlds where only I walk on the lawn are more similar to the actual world than those worlds where the lawn is stampeded by everyone in the neighborhood (or by everyone on Earth).

Given that much of which worlds are deemed relevant to the analysis is determined by the antecedent, there is no need to adjoin any indices to the variably strict conditional – there's only one. I have already touched on this in the previous section, when discussing the relevant antecedent worlds (sufficiently similar worlds) to the evaluation of the counterfactual and noting that Lewis intends there to be a fitting restriction of that kind for any antecedent of any strict conditional based on comparative similarity of worlds. I will also say more about this in the next section, once the formal model theory has been defined, but what has been illustrated by the above example should suffice for an intuitive outline of the rationale underlying the systems of spheres models for the variably strict conditional. Such a framework turns out to

be robust enough to express, not only Stalnaker's theory[80] of the counterfactual, but also other notable reformulations, and analogies in temporal and deontic logic, of which Lewis (1973) gives a comprehensive analysis. I won't discuss those general correspondences here – it will suffice to say that most theories of the counterfactual (as variably strict conditional), relevant to the current chapter, can be expressed in terms of Lewis' 'similarity sphere' semantics, which I define and discuss in detail in the next section.

2.2.3 Similarity Spheres semantics for counterfactuals[81]

In the previous couple of sections, we have shown how Lewis argues in favour of the view that counterfactuals are based on a strict conditional corresponding to an accessibility assignment determined by an *overall similarity of worlds*, and that no single strict conditional fashioned this way can adequately serve as a model for the counterfactual. I have also given an informal discussion in the previous section how Lewis' proposal to model the counterfactual as a variably strict conditional can be viewed as a direct modification of the intended variable sphere model. In this section I present the formal definition of the most basic model theory for the variably strict conditional, i.e. systems of spheres models, and give a critical overview of various conceptions of comparative similarity of worlds and the corresponding restrictions on the systems of spheres assignments, with attention to inferences that such restrictions validate. This overview will include a critical comparison of Lewis' and Stalnaker's theories, as well as a case for the system that I believe to be the most suitable.

Definition 2.13: A system of spheres $ is a function[82]

$$\$\colon W \longrightarrow \wp\big(\wp(W)\big)$$

that assigns to each $w \in W$ a set of subsets of W, and satisfies the following condition:

(S1) $\$_w$ is *nested*: for all $S, T \in \$_w$ either $S \subseteq T$ or $T \subseteq S$.

[80] Stalnaker (1968), Lewis (1973, §3.4). I'll say more about Stalnaker's account of the counterfactual in sections §2.2.7 and §2.2.8, when contrasting it with Lewis' account.

[81] To place Lewis' semantics in a broader semantic perspective, see (Pacuit 2017, §1.4.3) for an exposition of the fact that his systems of spheres are an instance of neighbourhood semantics. That is, sphere frames $(W, \$)$ are just neighbourhood frames (W, N), where $N(w)$ is interpreted as the neighbourhood around w, with the kinds of constraints placed on $\$_w$ that Lewis believed best represented the intended notion of overall similarity of worlds, applicable to his analysis of the counterfactual.

[82] Based on Lewis (1973, pp.13-14).

Theorem 2.4: For all $w \in W$, each $\$_w$ such that $\emptyset \in \$_w$ has the following properties:[83]

(S2) $\$_w$ is closed under finite unions: for every finite $\mathcal{S} \subseteq \$_w$, $\cup\mathcal{S} \in \$_w$.

(S3) $\$_w$ is closed under finite intersections: for every finite $\mathcal{S} \subseteq \$_w$ such that $\mathcal{S} \neq \emptyset$, $\cap\mathcal{S} \in \$_w$.

Both follow directly from S1.

Proof : Proof by induction on the size of $\mathcal{S}$. First, consider the trivial cases when $\mathcal{S} = \emptyset$, and the case when $\mathcal{S}$ is a singleton. When $\mathcal{S} = \emptyset$, then $\cup\mathcal{S} = \emptyset$, so $\cup\mathcal{S} \in \$_w$, by stipulation. When $\mathcal{S} = \{S\}$, then $\cup\mathcal{S} = \cap\mathcal{S} = S \in \$_w$. Now for the base case when $\mathcal{S} = \{S, T\}$. By S1, either $S \subseteq T$, in which case $\cup\mathcal{S} = T \in \$_w$ and $\cap\mathcal{S} = S \in \$_w$, or $T \subseteq S$ in which case $\cup\mathcal{S} = S \in \$_w$ and $\cap\mathcal{S} = T \in \$_w$.

Now, suppose that both $\cup\mathcal{S} \in \$_w$ and $\cap\mathcal{S} \in \$_w$ for $|\mathcal{S}| = k$, for some $k \in \mathbb{N}$. Next, suppose that $|\mathcal{S}| = k + 1$, and take any $\mathcal{T} \subseteq \mathcal{S}$ such that $|\mathcal{T}| = k$. Then by the induction hypothesis, we have both $\cup\mathcal{T} \in \$_w$ and $\cap\mathcal{T} \in \$_w$. Next consider $\mathcal{S}\backslash\mathcal{T} = \{S\}$. Then $S \in \$_w$. By S1, either $S \subseteq \cup\mathcal{T}$ in which case $\cup\mathcal{S} = \cup\mathcal{T} \in \$_w$, or $\cup\mathcal{T} \subseteq S$ in which case $\cup\mathcal{S} = S \in \$_w$. Similarly, either $S \subseteq \cap\mathcal{T}$ in which case $\cap\mathcal{S} = S \in \$_w$, or $\cap\mathcal{T} \subseteq S$ in which case $\cap\mathcal{S} = \cap\mathcal{T} \in \$_w$, as required. □

Note that from the definition of a system of spheres we have $\cup\$_w \subseteq W$. All worlds outside $\cup\$_w$ are to be regarded "as being all equally similar to *w*, and less similar to *w* than any world that the spheres reach" (Lewis, 1973, p.16).

Definition 2.14: The ordered pair $(W, \$)$ is *a frame based on a system of spheres*, where W is a set, and $\$$ is as given in *Definition 2.13*. For brevity, call such ordered pairs **S**-frames. On the intended interpretation, relevant to the semantics under consideration, the elements 0f W are *possible worlds.*

Having established the basic frame theory, we can now define the formal truth conditions for the extended language and give the basic model theory.

[83] Lewis (1973, p.15) argues that although somewhat unintuitive, it is technically convenient to leave the empty set in as a sphere around each centered world. However, it can be easily verified that the presence of the empty sphere has no effect at all on the difference to the truth conditions. I haven't stipulated S2 and S3, as Lewis did, but instead I have shown them to be a consequence of S1 (for arbitrary unions and nonempty intersections we need additional conditions, which I address in §2.2.7). Lewis stipulates S2 and S3 alongside S1, so the closure under unions implies $\emptyset \in \$_w$ for each w. Not stipulating S2, I had to stipulate $\emptyset \in \$_w$.

Definition 2.15: A system of spheres model (or **S**-model) is the triple $(W, \$, [\ ])$ such that:[84]

(1) $(W, \$)$ is an **S**-frame

(2) $[\ ]: For \longrightarrow \wp(W)$ assigns to each $A \in For$ a subset of W (worlds where A is true)[85]. Below is the recursive definition of $[\ ]$.

(3) $[\sim A] = W \backslash [A]$
$[A \wedge B] = [A] \cap [B]$
$[A \vee B] = [A] \cup [B]$
$[A \supset B] = [\sim A] \cup [B]$

(4) $w \in [\Box A]$ iff $[A] = W$
$w \in [\Diamond A]$ iff $[A] \cap W \neq \emptyset$
$w \in [A > B]$ iff $[A] \cap \cup \$_w = \emptyset$ or $\exists S \in \$_w (\emptyset \neq (S \cap [A]) \subseteq [B])$

Definition 2.16: **S**-validity

Let $\models_{\mathbf{S}} \subseteq \wp(For) \times For$, and define $\Sigma \models_{\mathbf{S}} A$ iff for all models $(W, \$, [\ ])$, and all $w \in W$, if $w \Vdash B$ for all $B \in \Sigma$, then $w \Vdash A$. That is, valid inference is defined as truth preservation at all worlds in all systems of spheres models. A formula $A \in For$ is said to be valid iff $\emptyset \models_{\mathbf{S}} A$. Call this logic **S**.

Definition 2.17: Entertainability

A formula $A \in For$ is said to be *entertainable* at a world $w \in W$ iff $[A] \cap \cup \$_w \neq \emptyset$.

2.2.4 Centering: strict *vs*. weak

Condition S1 leaves open a number of conceptions of comparative similarity of worlds. In particular, with regard to the location of w in $\$_w$. This is important, since each $\$_w$ for each world w has been set up for the purpose of evaluating counterfactuals at w. It should be noted that some additional constraint regarding the location of w in $\$_w$ is required to validate *Modus Ponens* for $>$, since **S** alone doesn't. That is:

Proposition 2.5: $p, p > q \not\models_{\mathbf{S}} q$

Proof : Consider the countermodel: $W = \{u, w\}, \$_w = \{\{u\}, \{u, w\}\}$, $\{u, w\} \subseteq [p]$, $u \in [q]$, $w \notin [q]$. It's clear that the system of spheres $\$_w$ satisfies S1-S3. Also, $w \in [p > q]$, since there is some sphere $S \in \$_w$, namely $\{u\}$, such that $\{u\} \cap [p] = \{u\} \subseteq [q] = \{u\}$. □

[84] Based on Lewis (1970, p.76), where he calls them β models.
[85] Thus $w \in [A]$ means '*A* is true at world *w*'.

Let us discuss two important restrictions on $, whose addition validates *Modus Ponens*, and critically examine any other implication that their addition would have for the theory. Lewis (in agreement with Stalnaker) insisted on the additional constraint, called the *centering* condition, motivated by the conception of comparative similarity of worlds whereby (i) any world *w* is as similar to itself as any other world is to it, and (ii) no other world is as similar to a world *w* as *w* is to itself.[86] This conception of comparative similarity motivates the following restriction on $:

Definition 2.18: Call the system $ *centered* on *w* iff for every $w \in W$: $\{w\} \in \$_w$.
Abbreviation: (C).

However, the system **S**+C validates inferring $A > B$ from assuming both the antecedent and the consequent, without regard for any connection, or lack thereof for that matter, between *A* and *B*.

Proposition 2.6: $A, B \vDash_{\mathbf{S+C}} A > B$
Proof : Suppose $w \in [A] \cap [B]$ for some $w \in W$ and $ satisfying S1 and C. Then $\{w\} \in \$_w$, and $\{w\} \cap [A] = \{w\} \subseteq [A] \cap [B] \subseteq [B]$. Hence $w \in [A > B]$, as required. □

But consider the following examples, which have the above form, yet seem intuitively invalid.

(1) Suppose I toss a fair coin and it lands heads, so the counterfactual 'If I tossed a fair coin, it would land heads' would be, erroneously evaluated as true, given that it is just as likely to land tails.

(2) Suppose I have scrambled eggs for breakfast, and (as usual) the Sun rises in the East, then 'Were I to have scrambled eggs for breakfast, the Sun would rise in the East' would be true, despite the absence of a causal, or any other relevant connection between the antecedent and consequent.

In general, most would agree that the truth of the counterfactual doesn't merely depend on the truth of the antecedent and consequent, but some connection obtaining between the two. Read (1995, p.94) suggested that if we interpreted similarity as *similarity in relevant respects*, then other worlds could tie in similarity with the actual world, thereby providing a

[86] Stalnaker (1968), Lewis (1973).

mechanism for invalidating the above inference.[87] This motivates a weaker centering condition than C – formulated, but not adopted by Lewis – which corresponds to the conception of comparative similarity of worlds whereby any world *w* is as similar to itself as any other world is to it. This conception of comparative similarity of worlds motivates the following restriction on $:

Definition 2.19: Call $\$_w$ *weakly centered* on *w* if and only if *w* belongs to every nonempty sphere around *w*, and there is at least one nonempty sphere around *w*, that is: $\$_w$ is *weakly centered* on *w* iff $w \in \bigcap(\$_w \setminus \emptyset)$ and $\exists S \in \$_w (S \neq \emptyset)$.
Abbreviation: (W).

The worlds that are allowed to tie in similarity to *w*, i.e. worlds in $\bigcap\$_w$ are interpreted as being as similar *in relevant respects* to *w* as *w* is to itself.

NOTE: C implies W, since $\{w\} \in \$_w$ and S1 implies $w \in \bigcap\$_w = \{w\}$. Hence, **S**+C is a proper extension of **S**+W.[88]

It turns out that adding W to S1, has the virtue of making the logic strong enough to validate *Modus Ponens*, but not as strong as to validate the aforementioned problematic inference that an advocate of C is committed to. That is, **S** +W has the following virtues:

Proposition 2.7: $A, A > B \vDash_{\mathbf{S}+\mathrm{W}} B$
Proof : Assume that $w \in [A]$ and $w \in [A > B]$ for some $w \in W$ according to some weakly centered system of spheres $. Now, $w \in [A]$ and $w \in [A > B]$, from hypothesis, which implies that there is some $S \in \$_w$ such that $S \cap [A] \subseteq [B]$. Given W and S1 we infer that $w \in \bigcap\$_w \subseteq T$ for any $T \in \$_w$, so in particular $w \in S$. Hence $w \in [B]$, as required. □

Proposition 2.8: $p, q \nvDash_{\mathbf{S}+\mathrm{W}} p > q$
Proof : Consider the following countermodel: $W = \{w, u\}, \$_w = \{\{w, u\}\}$, $w \in [p] \cap [q]$, $u \in [p]$, and $u \notin [q]$. It's clear that $\$_w$ satisfies S1 and W. But, $\{w, u\} = \{w, u\} \cap [p] \not\subseteq [q] = \{w\}$. □

This makes **S** +W the weakest system of that validates *Modus Ponens*.

[87] Nolan (1997, p.543) also endorses that approach.
[88] The system **S**+W is what Lewis (1973) calls **VW**, which is obtained from Lewis' preferred system **VC** (commonly referred to as **C1**) by replacing the *strict centering* condition with the *weak centering* condition, or equivalently removing the axiom $(A \wedge B) \supset (A > B)$ from the axiomatized version of **VC** (Lewis 1973, p.132).

2.2.5 Universality condition

We saw earlier, in §2.2.1.3 that on the comparative similarity of worlds approach to expressing *ceteris paribus* constrains not all antecedent-worlds matter in determining whether a counterfactual is true at some world – some worlds may be so *dissimilar* from the world at which the counterfactual is being considered that including them in the analysis would lead to a wrong result (recall the example with worlds where tailless kangaroos stay upright using crutches). This means that there may also be possible worlds that are so bizarre as to be left out of consideration to determining the truth of *any* counterfactual at the world of evaluation. Call such worlds *absolutely irrelevant*. In other words, given some $\$_w$, the question is whether there are any worlds that are precluded from determining the truth of any counterfactual at *w* due to their dissimilarity to *w*. Whether one wishes to completely rule out some worlds from the analysis (or not) motivates the formulation of another restriction on $:

Definition 2.20: Call the system $ *universal* iff for every $w \in W$: $\cup\$_w = W$.
Abbreviation: (U).

This condition offered in Lewis' general proposal is highly relevant to the semantics of counterpossibles. It concerns the limits of the accessibility relation, or equivalently, the limits of the similarity of worlds. Lewis leaves it open whether the union of all similarity spheres around some centered world should contain all possible worlds – it may not, and those worlds that are left out are to be interpreted as irrelevant to the evaluation of the counterfactual.

> If $\cup\$_i$ is the set of all worlds, for each *i*, I will call $ universal. If not, then I regard the worlds that the spheres around *i* do not reach – those that lie outside of the union of $\cup\$_i$ – as being all equally similar to *i*, and less similar to *i* than any world that the spheres reach. We will see that any such world will be left out of consideration in determining whether a counterfactual is true at *i*. It is as if, from the point of view of *i*, these remotest worlds were not possible at all. (Lewis 1973, p.16)

This is yet another example of how the character of Lewis' (1973) work makes his semantics not only amenable to the kind of extensions that shall be considered in this thesis, but also *highly suggestive* with regard to how one may actually proceed in doing so. From the above passage, it's reasonable to interpret impossible worlds in relation to the *universality*

restriction as those that are among the *absolutely irrelevant* worlds in evaluating a counterfactual at some possible world. I will return to this interpretation in *Chapter 5*.

2.2.6 The Limit Assumption

In the last few sections we have discussed a variety of conceptions of comparative similarity of worlds and saw how they translate to corresponding restrictions on $\$_w$ and what characteristic inference forms the respective systems validate. The assumptions underlying those various notions of comparative similarity included general limits on the similarity (centering conditions) and dissimilarity (universality condition) of worlds to w. The aspect of comparative similarity that we now turn to involves an *assumption* about similarity of antecedent-worlds to w, and the corresponding properties of $\$_w$ – namely whether for any entertainable antecedent A there should always exist the most similar A-worlds.

Because this assumption has consistently drawn the attention of philosophers, ever since its initial formulation, I'll devote a whole section to describing it and highlighting some of its key implications for comparative similarity theories of the counterfactual. Lewis (1973, §1.4) identifies a subtle and important property that one could very well assume comparative similarity of worlds to have, yet which isn't implied by S1. As we'll shortly see, the property in question corresponds to a generalization of the derived property S3 in *Theorem 2.4*.

If there are finitely many spheres $\$_w$ around some world w, then any non-empty set of those spheres has a smallest member (the same would hold if $\$_w$ was infinite and we only considered finite and non-empty subsets of $\$_w$). In particular, for any entertainable antecedent A, the set of all A-permitting spheres, i.e. the set $\{S \in \$_w : S \cap [A] \neq \emptyset\}$, has a smallest member, which is just the intersection of all A-permitting spheres: $\bigcap\{S \in \$_w : S \cap [A] \neq \emptyset\}$. This sphere is said to contain the *closest* (or *most similar*) A-worlds to w, i.e. all and only those A-worlds than which no other A-world is closer (more similar) to w. But if there are infinitely many spheres, there may not always exist smallest antecedent-permitting spheres around w, for every antecedent. If there are infinite sequences of smaller and smaller spheres without end, then there are sets of spheres without a smallest member (least element). In particular, for some world w and antecedent A there may not always be the smallest A-permitting sphere around w (containing the *most similar* A-worlds to w). To assume otherwise is to make the

Limit Assumption.[89] That is, it need not be the case that for every entertainable antecedent there always exists at least one most similar world.[90]

Definition 2.21: The *Limit Assumption*: for any $A \in For$, if $\cup\$_w \cap [A] \neq \emptyset$ then there exists a *smallest A*-permitting sphere (containing the most similar antecedent-worlds to *w*). We can formalize it as follows: for any $A \in For$ if $\cup\$_w \cap [A] \neq \emptyset$ then $\cap\{S \in \$_w : S \cap [A] \neq \emptyset\} \in \$_w$. Abbreviation: (LA)

With LA the truth conditions for the counterfactual could be greatly simplified:

Definition 2.22: Truth conditions for the counterfactual with the *Limit Assumption*
The counterfactual $A > B$ is true at world *w* (according to \$) if and only if

There is no *A*-permitting sphere in $\$_w$: $\cup\$_w \cap [A] \neq \emptyset$

or

B holds at every *A*-world in the smallest *A*-permitting sphere: $\cap\{S \in \$_w : S \cap [A] \neq \emptyset\} \subseteq [B]$

Or even simpler:

The counterfactual $A > B$ is true at world *w* (according to \$) if and only if

B holds at every *closest* (*most similar*) *A*-world: $\underline{[A]} \subseteq [B]$

(Where $\underline{[A]} := \cap\{S \in \$_w : S \cap [A] \neq \emptyset\} \cap [A]$ and no world outside of $\cup\$_w$ is *closest*).

Lewis observes that this assumption can't always be made, and consequently decides against adding it to the conditions that characterize \$. He argues that because there may be cases where LA is simply false, truth conditions from *Definition 2.22* would give the wrong analysis. Consider the following argument: I'm thinking (at this moment) of a line that is half a unit in length (in Euclidian 3-space). But let us suppose, counterfactually, that I thought (just then) of a line that was longer than a unit. That is, the counterfactual antecedent is 'I'm thinking of a line that is longer than a unit', abbreviated with '$1 < L$'. Next, we see that $\{S \in \$_@ : S \cap [1 < L] \neq \emptyset\}$ is the set of all $(1 < L)$-permitting spheres, i.e. spheres containing worlds where I'm thinking about a line that is longer than a unit. Now, here's a key step in Lewis' argument: the worlds where I'm thinking of a line that is 1½ units long are more similar to the actual world than worlds where I'm thinking of a line that is 2 units long, and

[89] (Lewis 1973, §1.4, pp.19-20)
[90] The Limit Assumption is explicit in Stalnaker's truth conditions for the counterfactual – which I discuss in the next section – so this argument can be viewed as being indirectly aimed at Stalnaker's analysis.

likewise the worlds where I'm thinking of a line that is 1¼ units long are more similar to the actual world than worlds where I'm thinking of a line that is 1½ units long, and so on, *ad infinitum*.[91] The shorter we make the line (above 1 unit), the more similar we make it to the length that I actually thought about, so presumably, the closer we come to the actual world.[92] But how long is the line that I'm thinking about in the *closest* (most similar) $(1 < L)$-worlds? The short answer is that there is *no such length*. [93] Since there is no smallest length that is greater than one unit, there are no worlds where I'm thinking about a line with such length, and consequently no smallest sphere containing all and only such worlds. Hence, $\cap\{S \in \$_w : S \cap [1 < L] \neq \emptyset\} = \emptyset$.[94]

On the truth conditions given in *Definition 2.22*, the counterfactual 'If I thought just then about a line that is longer than one unit, then…' would be vacuously and erroneously evaluated as true for any consequent. In particular, the following counterfactual would be erroneously evaluated as true: 'Had I been thinking about a line that is longer than one unit, then I would have been thinking about a line that is not longer than one unit'. But it would be evaluated vacuously as true not because the antecedent does not express an entertainable proposition (clearly it does, e.g. take worlds where I was thinking of the line being 1½ units long), but because there are no *closest* (most similar) accessible worlds where it is true.

There have been a number of replies to Lewis' argument against LA.[95] Some (Pollock 1976, Stalnaker 1980) demonstrate questionable implications of denying LA, others (Hájek 2014) argue that the implications about comparative similarity stemming from Lewis' formulation of the argument against LA commit him to the truth of clearly false counterfactuals. Another

---

[91] Lewis uses a different example – one involving a physical line (that is actually less than an inch long and the counterfactual supposition is that it is more than 1'' long), printed on a page, but admits that such examples aren't decisive, since given the printing process, there would be a limit on how long a line can be printed, being contingent on the printing process (e.g. the number of ink molecules on the page being finite). To avoid this, I have chosen a to exemplify an idealized (mathematical) line, which isn't subject to these kinds of physical restrictions. To make it unambiguous and precise, say that I've been *actually* thinking about the half-unit vector $0.5\boldsymbol{i}$ in the representation of Euclidian 3-space $\mathbb{R}^3$ (where $\boldsymbol{i}$ is the unit vector in the x-coordinate direction), and in the counterfactual supposition I'm thinking of some $l\boldsymbol{i}$ where $1 < l \in \mathbb{R}$.

[92] This part of Lewis' argument gets him in trouble (a point I'll return to shortly) and seems to be at odds with what he says earlier: "And so it goes; respects of similarity and difference trade off. If we try too hard for exact similarity to the actual world in one respect, we will get excessive differences in some other respect" (Lewis 1973, p.9).

[93] This follows from the property of real numbers that for any two real numbers $x < y$ there is a third real number z such that $x < z < y$.

[94] Based on Lewis (1973, §1.4, p.20).

[95] (Pollock 1976; Stalnaker 1980; Hájek 2014)

family of replies (Stalnaker 1980, Brogaard & Salerno 2013) involve appeal to a conception of comparative similarity – one that emphasizes the respects of similarity that are *relevant* to the context in which a counterfactual is being considered – which can diffuse the problematic consequence of accepting LA, identified by Lewis, and avoid the aforementioned issue raised by (Hájek 2014). I adopt a version of this conception of comparative similarity – one that is consistent with S1+W – and devote part of *Chapter 4* to motivating and defending it. For now, we turn to an example of a theory that has adopted the limit assumption.

2.2.7 Stalnaker's theory and conditional excluded middle

Having discussed various conceptions of comparative similarity of worlds in the previous sections, we have only hinted at the character of Stalnaker's theory. In this section we examine carefully the systems of spheres corresponding to the account of the conditional given by Stalnaker (1968, 1970), which, aside from the limit assumption places another significant restriction on $. It's yet another addition to the kinds of assumptions about comparative similarity that result in logics that are misaligned with the intended logic of counterfactuals.

I won't discuss the formal account of Stalnaker's (1968, 1970) original model theory, other than saying that he had originally based his semantics on Kripke frames, augmenting them with a world selection function $f: For \times W \longrightarrow W$ that selects for each antecedent-world pair (A, w) a single world $f(A, w)$ regarded as the *most similar* A-world to w. Lewis (1973, §3.4) shows how such selection-function models can be equivalently expressed in terms of his own similarity sphere models (given appropriate restrictions on $) – and that's the manner in which I choose to talk about Stalnaker's theory in this chapter, which after all is devoted to Lewis' semantics.[96]

Stalnaker's truth conditions for the counterfactual:

"Consider a possible world in which *A* is true, and which otherwise differs minimally from the actual world. '*If A, then B*' *is true (false) just in case B is true (false) in that possible world."*[97]

[96] For an account of conditional logics in terms of the selection functions, and a comprehensive comparison of Stalnaker's and Lewis' theories in terms of selection functions see (Priest 2008, §5) or (Lewis 1973, §2.7, §3.4). For a general discussion of the selection-function semantics for conditional logics see (Chellas 1975) or (Nute 1980, §3.2).

[97] (Stalnaker 1968, p.102)

Apparently, in addition to the *limit assumption* – that for any world *w* and entertainable antecedent *A* there is at least one *A*-world that is most similar to *w* – Stalnaker makes a stronger assumption, i.e. that there is a *unique* world like this.[98] It leads to an important difference between Lewis' and Stalnaker's accounts of the counterfactual. Adding this assumption to S1 validates *Conditional Excluded Middle* (*abbr*. CEM), i.e. '$(A > B) \vee (A > {\sim}B)$'.[99] Lewis calls CEM the "principal virtue and the principal vice" of Stalnaker's theory, presumably because although it may appeal to the intuitions of ordinary language users, nevertheless it's hardly true of the subjunctive conditional – as I'll shortly demonstrate *via* Lewis' argument against CEM.

Definition 2.23: *Stalnaker's uniqueness assumption*: for every world *w* and antecedent *A*, that is entertainable at *w*, there is a smallest *A*–admitting sphere around *w* containing exactly one *A*–world. Formally, $ is said to satisfy *Stalnaker's uniqueness assumption* if and only if for all $w \in W$ and $A \in For$:
if $\cup\$_w \cap [A] \neq \emptyset$, then $(\exists S \in \$_w)(\exists u \in W)(S = \cap\{T \in \$_w : T \cap [A] \neq \emptyset\} \wedge S \cap [A] = \{u\})$.
Abbreviation: (SA).[100]

Clearly, the *uniqueness* component of SA does all the work in the validation of CEM, since if for each antecedent there exists a smallest antecedent-permitting sphere containing only a single antecedent-world, then given that LEM is valid, the consequent of either disjunct must be either true or false at that world, thus making exactly one of the disjuncts true. Below is a proof.

Proposition 2.9: $\vDash_{\mathbf{S+SA}} (A > B) \vee (A > {\sim}B)$
Proof : Either $\cup\$_w \cap [A] = \emptyset$, or $\cup\$_w \cap [A] \neq \emptyset$. Both disjuncts are true for the vacuous case. If $\cup\$_w \cap [A] \neq \emptyset$, then SA allows us to infer that $\exists S \in \$_w \exists u \in W(S \cap [A] = \{u\})$. Either $u \in$

[98] In the relevant literature, this condition is sometimes referred to as *Stalnaker's assumption* (Lewis 1973), or *the uniqueness assumption* (Stalnaker 1980), or *Stalnaker's uniqueness assumption* (Nute 1980).
[99] (Lewis 1973, p.79).
[100] As for the treatment of impossible antecedents, Stalnaker also includes the *absurd world* where everything is the case and where all counterpossibles are evaluated, so SA also holds for impossible antecedents (Stalnaker 1970), (Lewis 1973, p.77). In *that* regard their theories are the same: on Lewis' account all counterpossibles come out vacuously true, by definition. They're also all true for Stalnaker since all consequents of counterpossibles are true at the absurd world. For good discussions of the differences between Stalnaker's and Lewis' similarity semantics for the counterfactual and the corresponding theories see (Lewis 1973, §3.4), (Read 1995, pp.82-95) and (Priest 2008, §5.7). I will return to the problem of vacuous counterpossibles in §2.3.

$[B]$ or $u \in [\sim B]$, by LEM, so, $\{u\} = S \cap [A] \subseteq [B] \ni u$ or $\{u\} = S \cap [A] \subseteq [\sim B] \ni u$. So, exactly one of the disjuncts must be true, by *Definition 2.15*. □

Proposition 2.10: $\nvDash_{\mathbf{S+C}} (p > q) \vee (p > \sim q)$[101]

Proof : Consider the countermodel: $W = \{w, u, v\}, \$_w = \{\{w\}, \{w, u, v\}\}$ and let $w \in [\sim p]$, and $u, v \in [p]$. The model clearly satisfies S1 and C. Also, let $u \in [q]$ and $v \in [\sim q]$. Now, $W \cap [p] \neq \emptyset$, since $\{w, u, v\} \cap [p] = \{u, v\} \neq \emptyset$, but neither $\{u, v\} \subseteq [\sim q]$ nor $\{u, v\} \subseteq [q]$. So, neither disjunct is true, by *Definition 2.15*, as required. □

The system corresponding to Stalnaker's theory is in fact a proper extension of all the systems we've discussed so far, since, if we recall he favors the stronger of the centering conditions C, so by adding SA, his preferred system **S**+C+SA is clearly a proper extension of Lewis' preferred system **S**+C. The logics corresponding to systems **S**+C and **S**+C+SA are known in the relevant literature as **C1** (or **VC**) and **C2**, respectively.[102]

Lewis admits that his theory was motivated by the observation that the whole appeal of CEM is due to ordinary language speakers rarely making the distinction between the external negation of a whole counterfactual, i.e. $\sim(A > B)$ and the same counterfactual with a negated consequent i.e. $A > \sim B$. This results in many ordinary language speakers choosing to reject the violation of CEM, because *prima facie* it appears to be a contradiction. Stalnaker's theory aligns with that intuition.

Lewis (1973) brings up Quine's (1950) example involving the renowned composers Bizet and Verdi, where their nationalities are counterfactually identified: as a matter of fact, Bizet was French, and Vivaldi was Italian. However, it's neither the case that if they were compatriots, Bizet would be Italian, nor is it the case that if they were compatriots, Vivaldi would be French. Nevertheless, certainly if they were compatriots, they'd be either French or Italian. That is, Lewis claims that the following conjunction is true, but given CEM it can't be, since it insists on the first or second conjunct being false: $\sim(A > B) \wedge \sim(A > \sim B) \wedge \big(A > (B \vee \sim B)\big)$.[103]

---

[101] Hence, $\nvDash_{\mathbf{S+W}} (A > B) \vee (A > \sim B)$.

[102] Lewis (1973, p.130), Nute (1980, p.53), Priest (2008, §5.7).

[103] (Lewis 1973, p.80). Note that $\sim(A > B) \wedge \sim(A > \sim B)$ is equivalent to the negation of CEM, by *De Morgan* laws.

Therefore, to allow for the aforementioned distinctions, CEM needs to be invalidated. On the similarity sphere semantics this is achieved by allowing similarity ties between worlds (as exemplified in the countermodel to CEM given earlier). That is, there is a tie in similarity to the actual world between a world w_F where the two composers are both French, and the world w_I where they're both Italian. Presumably a world w_G where they're both German, say, would be less similar to the actual world than w_I and w_F are.[104] So, if Bizet and Verdi were compatriots, then it seems that neither of the following is true:

- If Bizet and Verdi were compatriots, then they would both be Italian.
- If Bizet and Verdi were compatriots, then they would both be French (i.e. *not* Italian).

A more recent counterexample to CEM, inspired by Hájek (2014), is based on a probabilistic argument. Suppose I didn't toss a fair coin just now. But were I to toss that coin just now, would it land heads or tails? Surely it would land heads or tails, but it seems that both of the following would be false:

- If I tossed the coin just now, it would land Heads.
- If I tossed the coin just now, it would land Tails (i.e. *not* Heads).

On the assumption that we're dealing with a fair coin, claiming that either of the above is true would run in the face of the fact that we're dealing with probabilistic (stochastic) process. It appears that Lewis' original qualms with CEM have gathered wider acceptance among philosophers.[105] There are many other counterexamples, once the form of the error is understood, but I chose to highlight the one above since Hájek (2014) employs the stochastic character of such a coin-toss event in setting up a nondeterministic version of Fine's (1975) argument against Lewis' similarity account of the counterfactual. The point is that the failure of CEM makes Lewis' theory more resistant than Stalnaker's to such a family of objections.

2.3 Lewis' analysis of counterpossibles

One of the major drawbacks of Lewis' account of the counterfactual is that it evaluates all counterfactuals with impossible antecedents as true. Since $\bigcup\$_w \subseteq W$, for all models, and on the intended interpretation W is a set containing possible worlds, there are no spheres containing impossible worlds. In other words, impossible worlds are not entertainable. So,

[104] (Priest 2008, p.95)

[105] As it will be shown in *Chapter 4*, the failure of CEM serves as a good counter to some recent (Hájek 2014) objections to Lewis' general account.

considering the truth conditions for the counterfactual, then for all worlds $w \in W$ and any antecedent $A \in For$ of a counterfactual $A > B$ that expresses an impossible proposition $[A] \cap \mathrm{U\$}_w = \emptyset$. Consequently, for any $B \in For$, the counterfactual $A > B$ is evaluated as true – vacuously so, by Lewis' truth conditions for, as given in *Definition 2.15*. Hence, $\vDash_{\mathbf{S}} A > B$ holds for any B, whenever A expresses an impossible proposition, and *a fortiori* it holds for all extensions of **S**.

Proposition 2.11: $\vDash_{\mathbf{S}} A > B$ whenever A expresses an impossible proposition.
Proof : Antecedents expressing an impossibility are not entertainable, so by *Definition 2.15* any counterfactual with such an antecedent satisfies the vacuous condition at all worlds. □

However, the claim that all counterpossibles are vacuously true – and as such semantically uninformative, since their consequents make no contribution to the truth value – conflicts with our intuitions, as there appears to exist a plethora of cases where a counterpossible is clearly false or non-vacuously true. Consider the examples given below. In addition to their intuitive appeal (of non-vacuously meaningful truth values), a good case can be made for either their falsity or non-vacuous truth (to which I devote a large part of *Chapter 3*), yet Lewis' account falls short of offering the corresponding, adequate analysis.

(1) If Sally were to square the circle, then someone would have squared the circle.
(2) If Sally were to square the circle and I were to double the cube, then I would be Sally.
(3) If paraconsistent logic were correct, *ex contradictione quodlibet* would still be valid.

Consider the first two - whereas (1) seems true and non-vacuously so, it would seem odd to insist that (2) is false, yet a vacuous-account doesn't distinguish between their truth values. We can argue for the falsity of (3) as follows: since paraconsistent logics invalidate *ex contradictione quodlibet* by definition, the consequent of (3) runs contrary to the meaning of the antecedent.

To justify his position, Lewis gives the following argument, which to no lesser extent employs intuition than the examples of apparently false and non-vacuously true counterpossibles listed in the previous paragraph:[106]

[106] This isn't the only argument that Lewis gives in support of the vacuous analysis. There are others, but their adequate treatment is beyond the scope of this chapter. E.g. there's Lewis' famous *marvellous mountain* argument (Lewis 1986, p.7) against impossible worlds, which I discuss in the next chapter.

> Confronted by an antecedent that is not really an entertainable supposition, one may react by saying, with a shrug: If that were so, anything you like would be true!
>
> Further, it seems that a counterfactual in which the antecedent logically implies the consequent ought always to be true; and one sort of impossible antecedent, a self-contradictory one, logically implies any consequent. (Lewis 1973, p.24)

There are two justifications given by Lewis there. My reply is in line with the analysis given by Brogaard and Salerno (2013, pp.648-9). Even if we grant Lewis the claim that all counterfactuals with antecedents that aren't entertainable suppositions invoke triviality, it doesn't mean that (unless assumed) all impossibility-expressing antecedents fail to be entertainable suppositions. So apparently Lewis' first justification rests on the unwarranted assumption that all counterpossibles involve antecedents that are not entertainable.

In the second justification contained in the above passage, Lewis states that a counterfactual whose antecedent logically implies the consequent ought to always be true. That is, Lewis appeals to the validity of *conditional proof* (CP) for counterfactuals: if $A \vDash B$ then $\vDash A > B$. He then points to the validity of *ex contradictione quodlibet* (ECQ), i.e. $A \wedge {\sim}A \vDash B$ to suggest that at least counterpossibles with contradictory antecedents (or more generally, antecedents corresponding to a conjunction of formulae that form an inconsistent set) should be trivially true, i.e. $\vDash (A \wedge {\sim}A) > B$.

There are two things to note here: the first is that assuming CP and ECQ, as Lewis does, can only support the less general claim that only counterfactuals logically impossible antecedents are trivially true, rather than the general claim that all counterpossibles are trivially true – which would require a further assumption that every impossibility, of every kind, is equivalent to some contradiction. So even if we granted Lewis the right to make those assumptions in this context, then his argument justifies a vacuous analysis of only a narrow subclass of counterpossibles. The second thing to note is a point made by Brogaard and Salerno (2013) that assuming CP in this context is very much theory-laden, since anyone who is already convinced that there are false counterpossibles will hardly be persuaded by appeals to an inference to which false counterpossibles constitute a counterexample.[107] That is, CP is valid if and only if all counterpossibles are vacuously true. It appears that Lewis' above attempt at justifying vacuously true analysis of counterpossibles, amounts to little more than an expression of skepticism regarding the alternative.

[107] For example, Nolan's (1997) proposal is based on the rejection of CP for the counterfactual/counterpossible.

2.4 Summary

We have seen that conditional logics, to which Lewis' analysis of the counterfactual belongs, avoid commitment to some troublesome inference forms. We have also discussed Lewis' motivations for settling on a model of the counterfactual based on a variably strict conditional based on comparative similarity of worlds and argued for the system **S**+W as the one with the least number of questionable commitments. However even **S**+W gives a vacuous analysis of counterpossibles, which I believe is an inadequate analysis of the more general conception of the counterfactual. I've addressed some of Lewis' justifications for the vacuous analysis and shown them to be inconclusive or unconvincing. In *Chapter 3* I address some the metaphysical reasons that potentially explain Lewis' insistence on the vacuous analysis – that reply of mine is part of the general case I make for a possible and impossible worlds semantics as a very natural approach to a non-vacuous account of counterpossibles. The approach to developing a non-vacuous analysis of counterpossibles that I endorse and follow, is one which has gained a fair amount of interest in the last couple of decades, and which takes the Lewis-style account of the counterfactual (just presented) as a starting point, and introduces impossible worlds as a means of giving content to impossible antecedents.[108] An account of non-vacuous counterpossibles in terms of impossible world semantics, developed and given in *Chapter 5* proceeds in that manner, whereas *Chapter 4* develops a conception of comparative similarity that places emphasis on the respects of similarity that are *relevant* to the context in which a counterfactual is being considered.

[108] Nolan (1997); Mares (1997); Vander Laan (2004); Brogaard & Salerno (2013); Bjerring (2014); Berto (2014, 2017).

Chapter 3

David Lewis' *Marvelous Mountain* argument against impossible worlds.

'On the mountain both P and Q' is equivalent to 'On the mountain P, and on the mountain Q'; likewise 'On the mountain not P' is equivalent to 'Not: on the mountain P'; putting these together, the alleged truth 'On the mountain both P and not P' is equivalent to the overt contradiction 'On the mountain P, and not: on the mountain P'.

Lewis (1986)

According to the consistent theory of objects, the traditional and widespread idea that impossible objects are quite beyond logical reach […] depends upon the long-standing confusion between attributing inconsistent properties to an item (e.g. f and ~f) and inconsistently attributing properties to it (e.g. saying it has f and that it is not the case that it has f).

Routley (1980)

3.0 Introduction

This chapter contains a defense of the *extended argument from admissible paraphrase*, against Lewis' (1986) objection. The central feature of this defense is a refutation of Lewis' (1986) famous *'marvelous mountain' argument* which was set up as a *reductio ad absurdum* in Counterpart Theory (CT), and which amounts to a rejection of impossible individuals (impossibilia) and by extension of impossible worlds. To ensure the clarity of that refutation, an overview of the CT elements on which the success of the *reductio* hinges, will be summarized beforehand.[109]

The following defense of impossibilia in Lewis's theory – which ultimately reduces to pointing out that the commutative property (e.g. 'at world w: *not* A' being equivalent to '*it is not the case that* at world w: A') is illicitly ascribed to the restricting modifier and truth-functional connectives on the extended domain – parallels Meinong's defense of impossibilia

[109] The extended argument from admissible paraphrase, aka 'argument from ways' (Vander Laan 1997), is a quite common strategy in the literature that is employed in support of impossible worlds (Yagisawa 1988), (Vander Laan 1997), (Berto 2009), (Nolan 1997). See (Yagisawa 1988, p.183) for a modal realist account, (Vander Laan 1997, p.598) for an abstractionist account, and (Berto 2009, p.3) for a hybrid account, whereby "possible worlds are taken as concrete Lewisian worlds, and impossibilities are represented as set-theoretic constructions out of them".

in his theory of objects against Russell's charge that the theory violates LNC.[110] Both arguments defend the consistency of theories of impossibilia, by pointing to a key distinction between a wider and a narrower negation (and the corresponding rules), and pointing out that their scopes differ. In Meinong's case, the narrower, property negation 'there is an object that is *not* blue' is to be distinguished from sentential negation '*it's not the case that* there is an object that is blue' when impossible objects are considered. Thus, similarly to the foregoing denial of the derivability of a contradiction from the admission of world-bound individuals that instantiate contradictions, Meinong denies the derivability of a contradiction from the admission of object-bound contradictory properties. That is, given the incongruence of predicate/sentence negation scopes in the presence of impossible objects one cannot infer the contradiction '*a* is round and it is not the case that *a* is round' from '*a* being both round and not round'. Routley's summing up of Meinong's defense of a consistent theory of impossible objects places the following refutation of Lewis's *reductio* in historical perspective whilst highlighting the essential feature of contention.

> According to the consistent theory of objects, the traditional and widespread idea that impossible objects are quite beyond logical reach (that they violate the fundamental laws of logic, are not amenable to logical treatment, and hence cannot be proper subjects of logical investigation) depends upon the long-standing confusion between *attributing inconsistent properties to an* item (e.g. f and ~f) and *inconsistently attributing properties to it* (e.g. saying it has f and that it is not the case that it has f). Only in the second case would impossibilia be beyond the scope of consistent logic. It is now evident that this hoary confusion can be cleaned up by making appropriate negation scope distinction.
> (Routley 1980 p.89, my emphasis).

3.1 The extended argument from *admissible paraphrase* – a defense.

In the context of investigating the nature of the truth of subjunctives asserted by way of *reductio*, where he considers them as being instances of non-trivially true counterpossibles, Lewis (1973, p.24) for the sake of argument accurately envisages the overview character of

[110] Routley (1980,n p.89).

such an extension, which would proceed by positing impossible worlds, before quickly dismissing it, on grounds of it being founded on a *'confused fantasy'*.[111]

> [O]ne sometimes asserts counterfactuals by way of *reductio* in philosophy, mathematics, and even logic. These counterfactuals are asserted in argument, and must therefore be thought true; but their antecedents deny what are thought to be philosophical, mathematical, or even logical truths, and must therefore be thought not only false but impossible. These asserted counterphilosophicals, countermathematicals, and counterlogicals look like examples of vacuously true counterfactuals.
>
> There are other things they might be, however. They might not really be counterfactuals, but subjunctive conditionals of some other kind. More interesting, they might be non-vacuously true counterfactuals, understood in the way I have proposed; but so understood under the pretense that along with the *possible* possible worlds that differ from our world only in matters of contingent, empirical fact, there also are some *impossible* possible worlds that differ from our world in matters of philosophical, mathematical, and even logical truth. (The pretense need not be taken very seriously to explain what happens in conversation; it just might be that this part of our conversational practice is founded upon a confused fantasy.) (Lewis 1973, p.24)

What is contained in the phrase 'a confused fantasy'? It refers to the claim, which Lewis labels 'a pretense', that posits the existence of impossible worlds that differ from the actual one in matters of philosophical, mathematical and logical truth. Does Lewis label such an extension of his analysis a *fantasy* because there *are no ways* the world could not have turned out? But then we could use Lewis' own justification for possible worlds given in the form of his argument from admissible paraphrase, by merely extending it in support of impossible worlds, whilst maintaining its form.[112] So, if Lewis' argument from admissible paraphrase for possible worlds is sound, then the soundness of the extended justification is only conditioned

[111] I take that what Lewis means by subjunctives asserted by way of reductio, are subjunctives of the form where the antecedent is the hypothesis of the reductio argument and the consequent the absurd conclusion derived. That is given the reductio 'HYP… ⊥,' Lewis has in mind the subjunctive 'HYP > ⊥'.

[112] The argument appears throughout Lewis' work. The versions of the argument I'm primarily relying on are taken from Lewis (1973, p.84) and Lewis (1986, p.2). Vander Laan (1997, §3) refers to it as 'the argument from ways'.

on whether one accepts the rather uncontroversial premise that not everything is possible.[113] Here is the original version of Lewis' argument:

> Ordinary language permits the paraphrase: there are many ways things could have been besides the way they actually are. On the face of it, this sentence is an existential quantification. It says that there exist many entities of a certain description, to wit 'ways things could have been'. I believe that things could have been different in countless ways; I believe permissible paraphrases of what I believe; taking the paraphrase at its face value, I therefore believe in the existence of entities that might be called 'ways things could have been'. I prefer to call them 'possible worlds'. (Lewis 1973, p.84)

To reiterate, if we accept this argument, then why should we not accept the following argument that there are impossible worlds?[114] The extended argument is a conditional thesis: if the paraphrase argument justifies belief in possible worlds, as ways things could have been, then by parity of reasoning, the same form of the argument justifies belief in impossible worlds, as ways things could not have been.[115] Being a conditional thesis, the full parity of reasoning argument can also be viewed as a *reductio* of genuine realism, directed to those who commit to concrete *possible* worlds only.[116]

The conditional argument first appears in Naylor (1986, pp.28-29) where it is presented in a way that could be interpreted as a direct *reductio* of genuine realism. It has also been taken up by Yagisawa (1988, p.183) where it serves as a lynchpin thesis in the conditional endorsement of extended modal realism. However, Yagisawa leaves it up to the reader whether the conditional thesis is to be taken as serving the *modus ponens* argument endorsing concrete impossible worlds, or the *modus tollens* arguments that would effectively echo Naylor's (1986) intended *reductio* of Lewis's justification of possible worlds, the soundness of which is premised on a consensus that impossible worlds do indeed lead to trouble. In Naylor' (1986) note to Lewis, the implication seems to be that a conclusion to the effect that the argument can be shown to speak equally in favour of impossible worlds is trouble enough. This is implicit since Naylor appears to expect the extended conclusion to speak for

[113] Mortenson (1989), is the rare exception to that view.
[114] Naylor (1986, p.29)
[115] Divers (2002, p.68)
[116] In fact the extended argument can be viewed as a reductio of Lewis's theory of genuine possible worlds (Yagisawa 1988), (Divers 2002).

itself without any further comment, since he doesn't bother to make one. But this is not really enough without an independent reason as to what is troublesome about positing impossible worlds. Moreover, Skyrms (1976, p.326) had already warned against caricaturizing Lewis's argument in a way that *ignores* the included proviso that taking the paraphrase at face value is only justified insofar as it doesn't lead to trouble.

> I do not make it an inviolable principle to take seeming existential quantifications in ordinary language at their face value. But I do recognize a presumption in favor of taking sentences at their face value, unless (1) taking them at face value is known to lead to trouble, and (2) taking them some other way is known not to. Lewis (1973, p.84)

But naturally there is no consensus as to what 'trouble' exactly amounts to. There is however a predicament that a classicist would wish to avoid, which would be taken as sufficient grounds to reject the extended argument, without abandoning the original one – namely, theoretical inconsistency. As it will be shown, Lewis gives an argument precisely to that effect, albeit a bad one. For him this is reason enough to reject the parity of reasoning argument, but it is a reason premised on a misunderstanding highlighted in the fragment from Routley, which I demonstrate in the sext section. Consequently, I conclude that the extended argument from admissible paraphrase in favour of impossible worlds is safe from the charge of running into the kind of trouble that Lewis believes it does.

To appreciate Lewis's reasons for banishing impossibilia from his ontology, it is necessary to give at least a rudimentary outline of those key elements of his metaphysics that are the primary suspects in being responsible for this "impossibilia phobia".

3.2 The status of objects in possibilist realism: an outline of *Genuine Realism* (GR).

Possible worlds exist *simpliciter* according to *genuine realism*, henceforth abbreviated with GR. They are spatiotemporally and causally isolated individuals, made up of mereological sums (fusions) of their parts. All parts of a world *w* stand in some spatiotemporal relation to each other, and if anything is spatiotemporally related to any part of *w*, then it is also a part of

w. Modal idioms reduce to unrestricted existential quantification over that which exists *simpliciter*, i.e. worlds, and objects that stand in a parthood relation to them.[117]

> You might say that strictly speaking, only this-worldly things *really* exist; and I am ready enough to agree; but on my view this 'strict' speaking is *restricted* speaking, on a par with saying that all the beer is in the fridge and ignoring most of the beer there is. When we quantify over less than all there is, we leave out things that (unrestrictedly speaking) exist *simpliciter*. Lewis (1986, p.3).

> The *genuine realist* (GR) takes unrestricted first-order existential quantification to range over a domain of individuals among which only some actually exist. Divers (2002, p.21)

GR possible worlds are just as real as <u>*our* world</u>. That is, actuality is *indexical*, i.e. relative to the world where 'this world' is uttered.[118] So the sentence token 'our world' in the previous sentence renders the world where it has been written down (this(!) world) as actual.

> When such expressions occur with their primary sense, their function is straightforwardly token-reflexive – that is, in any world w (in any sentential context of any sentence token) the expression introduces the world in which the token is uttered. To call a world 'actual' in this primary sense, is like referring to this place as 'here', or to this time as 'now' or to oneself as 'I'.
> (Divers 2002, p.44)

Modality reduces to quantification over worlds and their parts. This isn't unique to GR. However, GR has the virtue of a single, metaphysically undifferentiated domain of quantification – all worlds, and their parts are of the *same ontological kind*, i.e. differing not in kind, but only what goes on at them.[119] This allows GR to make use of unrestricted existential quantification – existence of any ***x*** is *existence simpliciter*; either ***x*** is itself a world, or it is *part* of some world.

---

[117] There is the odd feature of GR that worlds are taken to be individuals, *not* sets, even though they seem to contain stuff just as sets would. That is, in GR worlds don't *contain* their respective parts, but rather *consist* of them, or more precisely *stand in a parthood relation* to them. For details, see Divers (2002, pp.45-46).
[118] Ref. Lewis' indexical theory of actuality expressed in Lewis (1970) *Anselm and Actuality*.
[119] Lewis (1973, p.85).

> The domain of quantification is to contain every possible world and everything in every world. Lewis (1968, p.114)

> If asked what sort of thing [possible worlds] are […] I can only ask [my questioner] to admit that he knows what sort of thing our actual world is, and then explain that other worlds are more things of *that* sort, differing not in kind but only what goes on at them. Lewis (1973, p.85)

> 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.
> Lewis (1986, p.2)

> This means that modal extensionalism has a metaphysical edge over intensionalism, which postulates an extra *sui generis* metaphysical kind. And that is a significant philosophical virtue of modal extensionalism.[120]
> Yagisawa (1988, p.178)

This is a good place to introduce the key distinction (one that will be highly relevant in this chapter) between GR and *actualist realism* (aka *actualist representationism*) (AR), where according to the latter only the actual world exists, and non-actual possible worlds are mere abstractions (conceived of in one way or another).

> Broadly, GR conceives of the possible worlds as a vast plurality of non-actual, concrete things while AR conceives of the possible worlds as a vast plurality of actual, abstract things. (Divers 2002, p.22)

GR gives an *eliminative extensionalist* account of intension, modality in particular (Yagisawa calls it *modal extensionalism*), much in the same way as *actualist extensionalism* eliminates intensional notions such as properties and relations by replacing them with some extensional entities found in the actual world.[121]
Lewis (1965) proposes a translation of the language of quantified modal logic (QML) into a first-order logical theory, which he coins *Counterpart Theory* (CT). "Conceptually, GR intends CT as an element in the non-modal analysis of modal concepts. Semantically, GR

[120] Yagisawa (1988) calls Realist Possibilism and its (proposed therein) extension, which admits impossible worlds, Realist Impossibilism by what he considers to be their essential feature, i.e. *modal extensionalism*.
[121] Yagisawa (1988, pp.177-178).

holds CT essential to capturing the expressive capacity of modal English and CT may serve as a metalanguage that articulates a PW-semantics for QML" (Divers, 2002, p.123).

> Instead of formalizing our modal discourse by means of modal operators, we could follow our usual practice. We could stick to our standard logic (quantification theory with identity and without ineliminable singular terms) and provide it with predicates and a domain of quantification suited to the topic of modality. That done, certain expressions are available which take the place of modal operators. The new predicates required, together with postulates on them, constitute the system I call *Counterpart Theory*. Lewis (1968)

For the purposes of this chapter the following characterization will suffice. I present below those elements of CT that will be relevant to our discussion. The primitive predicates of counterpart theory relevant to our discussion are:

Wx (x is a possible world)
Ixy (x is in possible world y)

The domain of quantification is to contain every possible world and everything in every world. The primitives are to be understood according to their English readings and the following postulates:

P1:	$\forall x \forall y(Ixy \rightarrow Wy)$	Nothing is in anything except a world.
P7:	$\exists x\big(Wx \,\&\, \forall y(Iyx \leftrightarrow Ay)\big)$	Some world contains all and only actual things.

Key relevant GR predicates used by Divers (2002) are defined as follows:

$W\alpha$ (α is a world)
$P\alpha\beta$ (α is a part of β)

Consider the following absolute alethic *de dicto* possibility and its translation from English (1), to its *admissible paraphrase* with a direct reference to possible worlds (2), and finally its explicit GR interpretation (3), and the explicit CT translation in (4).[122]

(1) There may have been blue swans.
(2) There is a possible world at which there are blue swans.

[122] The translations (1)-(3) are given by Divers (2002, p.43), and the latter (4)-(6) being the original formulation and translation given by Lewis (1968).

(3) $\exists x \exists y[Wx \,\&\, Pyx \,\&\, Sy \,\&\, By]$.

(4) $\exists y(Wy \,\&\, \exists x(Ixy \,\&\, Sx \,\&\, Bx))$.[123]

Note the restricting modifier $Wy \,\&\, \exists x(Ixy \,\&\, \dots)$ in the scope of the outermost quantifier in (4). Its purpose is to restrict the domains of quantifiers that appear in its scope. Universal quantifiers would be restricted analogously. Below is the standard translation of the modal formulas, which is not unique to GR. What is unique to GR, as the *Black/blue Swans* example (below) from Lewis' Counterpart Theory (CT) will show, is the manner in which the world-indexed (world-restricted) formulae are treated.

(5) $\Box\varphi$ translates to $\forall y(Wy \supset \varphi^y)$

(6) $\Diamond\varphi$ translates to $\exists y(Wy \,\&\, \varphi^y)$

To form the world-indexed (world-restricted) sentence φ^y(φ holds in world y), the range of each quantifier appearing in φ is restricted to the world y. That is, the quantification '$\forall x$' appearing in the formula φ is replaced by $\forall x(Ixy \supset \cdots)$, and similarly, '$\exists x$' in φ is replaced by $\exists x(Ixy \,\&\, \dots)$, as we've already seen in (4).[124] Recall that existence at some possible world is the *same kind* of existence as at the actual world, so the usual quantifiers (albeit restricted to the given world) are used.

> [T]he phrase 'at *w*' […] works mainly by restricting the domains of quantifiers in their scope, in much the same way that restricting modifier 'in Australia' does. In Australia, all swans are black – all swans are indeed black, if we ignore everything not in Australia; quantifying only over things in Australia, all swans are black. At some strange world *w*, all swans are blue – all swans are indeed blue, if we ignore everything not part of the world *w*; quantifying only over things that are part of *w*, all swans are blue. (Lewis 1986, p.5)

The above, brief characterization of the role of such modifiers was primarily intended to aid a clear appreciation of Lewis' argument for rejecting the admission of impossible worlds into his modal realist ontology. Let's now turn to the analysis of his argument.

---

[123] To be precise, (4) is the Counterpart Theory (CT) translation of the quantified modal logic *de dicto* expression '$\Diamond\exists x Fx$'. Lewis (1968, p.118). Also, Lewis (1968) uses and Divers (2002) uses for propositional conjunction. I leave in the distinct symbols for fidelity's sake.

[124] Lewis (1968, p.118).

3.3 Trouble in paradise? – impossibilia and CT.[125]

The 'marvellous mountain' argument against impossible worlds.

Lewis's rejection of impossibilia, and consequently impossible worlds is, more precisely, the rejection of the existence of genuine world-bound individuals that instantiate impossibilities.[126] Lewis goes on to argue that the admission of inconsistent objects in a GR framework (hence impossible from a classical perspective) leads to a literal contradiction, and that is *trouble enough* for him to turn down impossible worlds, since he endorses ECQ (and needless to say he is not a trivialist). That is, Lewis argues that CT, which is a GR theory, is rendered inconsistent on the assumption of the existence of some classical impossibilia (objects with contradictory properties). I'll show in the next section that this argument is *unsound* since Lewis assumes that classically impossible worlds have classical properties – in particular, he assumes that the 'at world' restricting modifier commutes with truth-functional connectives for formulae that hold at inconsistent worlds.[127]

> [S]uppose travellers told of a place in this world – a marvellous mountain, far away in the bush – where contradictions are true. Allegedly we have truths of the form 'On the mountain both P and not P'. But if 'on the mountain' is a restricting modifier, which works by limiting domains of implicit and explicit quantification to a certain part of all that there is, then it has no effect on the truth-functional connectives. Then the order of modifier and connectives makes no difference. So 'On the mountain both P and Q' is equivalent to 'On the mountain P, and on the mountain Q'; likewise 'On the mountain not P' is equivalent to 'Not: on the mountain P'; putting these together, the alleged truth 'On the mountain both P and not P' is equivalent to the overt contradiction 'On the mountain P, and not: on the mountain P'. That is, there is no difference between a contradiction within the scope of the modifier and a plain contradiction that has the modifier within it. So to tell the alleged truth about the marvellously contradictory things that happen on the mountain is no different from contradicting yourself. But there is

[125] Lewis refers to modal realism as 'A Philosopher's Paradise'. Also, that's the title of Ch.1 in Lewis (1986). Here 'trouble' is taken to be anything that would be unacceptable to Lewis – in this case an inconsistent theory.

[126] (Divers 2002, p.67). Because in CT all individuals are by definition part of some world, so "individuals" whose parts are not spatiotemporally related, i.e. cross world "individuals" do not count as possibilia. I follow Divers in interpreting Lewis' marvellous mountain as a world-bound impossible individual.

[127] Commutativity here is understood as functional commutativity, where one function is the 'at w:' modifier and the other functions are the truth functions, e.g. negation, appearing in the scope of the modifier 'at w: not A'. On the assumption of commutativity 'at w: not A' is equivalent to 'not at w: A'. I'll make this more precise later.

> no subject matter, however marvellous, about which you can tell the truth by contradicting yourself. (Lewis 1986, p.7, f.1)

3.4 No trouble: the 'marvellous mountain' argument is *unsound.*

Lewis assumes that all restricting modifiers such as 'In Australia', 'On a mountain far away' or 'at world w' (abbreviated 'at w') commute with respect to truth-functional connectives in regimes other than classical ones.

The following objection to Lewis' 'marvellous mountain' argument, which he gave against the existence of inconsistent worlds, is not widely proposed among rival theories to GR.[128] There is a tendency among authors working in impossible world semantics, especially those who endorse non GR approaches to possible and impossible worlds to see the force of Lewis' argument stemming from the *metaphysical* aspect of its concretism – of the assumed nature of possible worlds in CT – as being a sufficient condition (and the key culprit), and only few take issue with the commutativity of the 'at w' modifier with the truth-functional connectives. That is, it seems to me that some authors are a little too quick to accept Lewis' derivation of the literal contradiction, and blame its success on the GR ontology — an alleged shortcoming of GR they like to point out as the fulcrum of that derivation.[129] For those authors, granting Lewis this apparently absurd consequence seems just a little too convenient, and indeed often serves as their cue to endorse less committal ontologies.[130] For example, Nolan (1997, p.541) thinks that Lewis' emphatic rejection of impossible worlds largely flows from what he takes them and their parts to be – namely as existing simpliciter. This, Nolan identifies as reason enough to derive a contradiction from positing objects with inconsistent properties.

> Extending this approach to impossible objects produces literal impossibilities, it seems: if the impossibilium corresponding to the blue swan-and-not-a-swan is literally a swan and is literally not a swan, then a contradiction is literally true. (Nolan 1997, p.541)

---

[128] A similar version of the argument can be found in Kiourti (2010, Ch. IV, §4.41). Mares (2004, pp.84-87) sees the problem with this property in the case of negation, which as a matter of fact *is* the key culprit. Lycan (1994, pp.39-41) believes that Lewis' argument fails, observing that a truly inclusive quantifier would require the invalidation of the entailment: at w: ~P entails ~(at w: P).

[129] E.g. (Nolan 1997, p.541), (Jago 2012, p.64), (Vander Laan 1997, p.606).

[130] Kiourti (2010, p.102) also makes this observation.

But this analysis takes it for granted that Lewis' *reductio* argument is sound. That is, Nolan takes it for granted that it's fine to go from *there being some individual that is both a swan and not a swan,* to *it being the case that the said individual is a swan and it not being the case that it is a swan*. However, the applicability of that inference doesn't rest on the *metaphysical* nature of the objects whose existence is being posited, but on the *logica*l principles thought as correspondingly fitting the metaphysical view, and consequently employed in the analysis. Vander Laan (1997, p.606) thinks that Lewis' reasons for rejecting impossibilia stem from his concretism and his insistence to take 'at w' as a restricting modifier, i.e. as merely restricting quantification over concrete worlds that are said to exist much in the same way as our world.

> Lewis goes on to say that 'at so-and-so world' is indeed a restricting modifier, unlike 'in such-and-such story', since worlds are like the actual world, not like stories. It is this last point that is of interest here. Lewis's reasons for rejecting impossible worlds stem from his concretism, that is, his view that worlds are concrete objects much like us and our surroundings. (Vander Laan 1997, p.606)

But this analysis assumes as correct Lewis' analysis of concrete impossibilia. Lewis' rejection of classically impossible worlds does not rely on the *metaphysical* nature of the objects in the quantifier's scope, but rather on the *logical* assumptions he makes about them, which are embodied by the posited properties of the 'at w' modifier. Vander Laan then suggests an abstractionist approach that treats worlds more like stories rather than concrete objects, i.e. where 'at so-and-so world' modifier is intended along the lines of 'according to such-and-such story'. Such an approach, he observes, would avoid the problem of ending up in contradiction, since stories, unlike worlds, need not be maximal nor consistent.

How should we read *'on the mountain"*? Let's recall that such modifiers act by restricting the quantifiers in their scope to the domain of a single world, much in the same way as the 'in Australia' restricting modifier restricts all talk to that which exists in Australia. So, all swans are indeed black if we restrict our discourse only to what exists in Australia. Lewis assumes the existence of a domain—the marvellous mountain—where contradictions are true, but for simplicity we can speak of maximal domains, i.e. worlds, where contradictions are true. The objection to Lewis doesn't hinge in any way on such domain generalisation, but rather will simplify the discussion – it will be a lot simpler to speak of worlds than their subdomains.

The premise that assumes the existence of a mountain such that 'On the mountain P and not P', amounts to assuming the existence of an inconsistent world, since according to Lewis's version of GR anything that exists must do so at some world.[131] So were anything to exist 'on the mountain', it would exist 'on the mountain at some world'. So, since we're considering the marvellous goings-on 'on the mountain' then *a fortiori* we're considering those goings-on at the world whose part is the marvellous mountain.

The key point to appreciate here is that since Lewis insists that the modifier and truth functional connectives commute in general, then it follows that this commutativity holds for the particular case when the modifier restricts quantification to the entire domain of *some* world. This shift of domain does somewhat reduce the generality of the original argument, but the refutation works equally well. That is, given that individuals are world bound in CT, and assuming in line with Divers that the marvellous mountain is intended as a world-bound impossibilium, extending the scope of the 'on the mountain' restricting quantifier to the world *w* of which the mountain is a mereological part of, and employing 'at *w*' instead, will not result in an omission of what happens on the mountain relative to the *actual* world. The point is that if a special case (here, about certain spatiotemporally related mereological sums of individuals) of a general claim (here, about any individual) is refuted, then so is the general claim. And the choice to lay out the refutation focusing on the special case is motivated only by clarity and simplicity of presentation.

One further thing to note is that given some world *w* (say the actual world) and some mereological subdomain of it (of the actual world *X*, say 'Australia in the year 2015') the truth of 'at *X*: *φ*' implies the truth of 'at *w*: *φ*' if and only if *φ* expresses an existential proposition, and the converse is true if *φ* expresses a universal proposition, for any *X* that is part of *w*. To see this, observe that *'in Australia in 2015, there are wombats'* implies *'at the actual world (i.e. in some spatiotemporal location), there are wombats'*, but *'in Australia in 2015 all swans are black'* doesn't imply that *'actually, all swans are black'*. Conversely actually *'every human is a mammal'* implies *'all humans in Australia are mammals'*, but *'there exist giant black holes'* doesn't imply that *'in Australia there are black holes'*. In my shift of domain (to worlds, from mountains) in the present refutation, I have taken care to avoid any possible issues that could arise due to the negligence of those relationships.

[131] In (Lewis 1968, p.114) axiom P1.

Lewis wants us to accept a certain property of the 'at w' modifier – namely that it commutes with the truth-functional connectives. Let's denote this alleged property of the modifier with MC, for modifier *commutativity* with truth-functional connectives:

(MC) For *any* domain X, and corresponding restricting modifier η_X, truth-functional connective/operator f, and sentence(s) φ: [132]

$$(\eta_X \circ f)(\varphi) \text{ iff } (f \circ \eta_X)(\varphi)$$

For clarification (examples), see the special cases of MC defined on the next page – namely MCC and MCN. Naturally, in this general definition worlds are just special kinds of domains. So, it is clear that if X is a subdomain of some world and is equivalent to a proposition with the I or O form, then the truth of $\eta_X(\varphi)$ entails the truth of $\eta_w(\varphi)$, and the converse is true when φ is either of the A or E form.[133]

What reasons does Lewis give in support of that property of the modifier? Well, given that the modifier restricts the domain of all quantifiers within its scope to one possible world, and given that worlds are (by definition of Lewis' CT) just mereological sums of their parts, what occurs at any given world, or region of some world, is amenable to purely mereological (extensional analysis). That is, saying 'in the box there's a green marble *and* a red marble' is the same as saying that 'in the box there's a green marble' *and* 'in the box there's a red marble'. That is 'at w: *A and B*' should be equivalent to 'at w: *A*' *and* 'at w: *B*', which indeed most people will accept as intuitively reasonable (by MC).

So far so (almost) good, it would seem. But Lewis then swiftly proceeds to harness our intuitions further (as if that was a completely seamless, immediate step), whilst we're still under the spell of the apparently innocuous[134] nature of the instance of MC just presented, and asks us to accept as equally innocuous to have MC extended to apply to yet another truth-functional connective – namely negation, i.e. 'at *w*: *not A*' and '*not* at *w*: *A*' being equivalent – in other words, negation and the restricting modifier *commute*. That is, we're supposed to accept this broader applicability as *equally* unproblematic, and not only when applied to possible worlds, but apparently to impossible ones as well, since it is explicitly applied to the

[132] Where ∘ denotes the function-composition relation. When *f* is a unary operator (e.g. negation) then it acts on a single sentence, else if *f* is a dyadic connective (e.g. conjunction) acts on a pair of sentences. For clarification (examples), see the special cases MCC and MCN, further below.

[133] **A**: Every S is P. **E**: No S is P. **I**: Some S is P. **O**: Some S is not P.

[134] Varzi (1997) gives a coherent account of non-adjunctive worlds.

'on the mountain' restricting modifier, and the mountain has been stipulated to be an inconsistent regime. Let's denote these instances of MC as follows:

(MCC) '<u>at *w*</u>: *A* and *B*' iff '<u>at *w*</u>: *A* and <u>at *w*</u>: *B*'.

(MCN) '<u>at *w*</u>: not *A*' iff 'not <u>at *w*</u>: *A*'.

In fact, the argument hinges on MCN holding. But it cannot hold.[135] To be sure, MC holds for *classical* worlds. That is, the commutativity of the 'at w' restricting modifier holds for *classical* truth-functional connectives, i.e. whose properties do not violate the principles of classical logic, but such principles are *not* guaranteed to obtain at non-classical worlds and the corresponding commutativity principle fails in such situations. Hoping that the reader's intuitions about classical mereology can work in his favour, in a way that they may be guided by the common-sense analogy: 'in the box there *is no* green marble' therefore *it's not the case that* 'in the box there *is* a green marble'. But affirming this purported property of the restricting modifier, as I will show, makes the unwarranted assumption that it ought to hold for inconsistent worlds, which ultimately boils down to wrongly assuming that inconsistent objects, and worlds of which they are a part, are to be analysed classically; in particular, that LNC ought to hold there. That is, MCN presupposes LNC, i.e. if LNC fails so does MCN.[136]

It will be of benefit to have a clear outline of the essential elements of Lewis's argument and its immediate consequences.

(Hyp) There is a world *w* such that <u>at *w*</u>: *A* and *not A*.

(P1) MCC and MCN are true of the domain restricting modifier for all worlds.

(C1) There is a world *w* such that <u>at *w*</u>: *A* and *not* <u>at *w*</u>: *A*. (Hyp.+P1)

(P2) There is no subject matter whereby one can tell the truth by contradicting themselves (LNC).

(C2) Therefore, there is no such *w*.

---

[135] Kiourti (2010, pp.116-119) raises similar objections to accepting MCN and consequently follows a very similar line of reasoning to mine in demonstrating what this assumption amounts to. My conclusion differs by virtue of how I have formulated the objection – Kiourti concludes that Lewis *begs the question* against the hypothesis of concrete impossibilia, where I conclude that the argument is unsound on grounds of the falsity of hypothesis concerning the properties of the 'at *w*' modifier, where *w* needn't be a classical world.

[136] Given that Lewis is committed to a classical truth conditional theory of meaning, his adherence to MCN is understandable. After all, denying MCN may require a denial of the classical truth conditional theory of meaning or denying that the connective '¬' means *not* (classical negation) at impossible worlds. However, these alternatives cannot be ignored without the risk of begging the question. For example, one may reject MCN by adopting an information theoretic theory of meaning, e.g. see (Mares 1997, 2004) – a theory which requires dialethism and a paraconsistent view of negation, both of which Lewis rejects emphatically.

(C3) Therefore there *are no* inconsistent worlds.[137] (RAA)

(C4) There are no impossible worlds.

Why assume that MCN ought to always hold for *all* worlds, both possible *and* impossible?[138] Let's think of circumstances that would violate MCN. Consider a world *w** that contains some genuinely inconsistent object *a*. That is, let us assume that 'at *w**: *Pa* and *not Pa*' for some property *P*. Now, given that *Pa* is a truth value glut (both true and false, but in particular *true*) at *w**, we *should analyse it accordingly*, which means we have no justified recourse to MCN, which presupposes LNC, *precisely because* the existence of *a* at *w** is a counterexample to LNC! That is, we cannot infer '*not* at *w**: *Pa*' from 'at *w**: *not Pa*', precisely because both *Pa* and *not Pa* hold at *w*.[139]

In other words, the assumption of the universality of MCN, rests on the erroneous supressed assumption that LNC ought to hold for inconsistent worlds in much the same way as it does for consistent ones, like boxes and fridges, which we are familiar with. But we are *not* considering consistent worlds, so appeal to MCN is unjustified, *precisely* because *w** is an inconsistent world. We will run into literal contradictions, i.e. 'at *w Pa*' and '*not* at *w Pa*' being true *simpliciter*, only if we assume the existence of truth value gluts, and then proceed to analyse them classically by unwarranted appeal to MCN. This is what Lewis in fact does. By refraining from any such questionable steps, we don't end up in explicit contradiction. So, although *w** is a genuinely inconsistent world, the theory remains consistent. That is, assuming the existence of *w**, doesn't lead to a contradiction *simpliciter*. But if the only reason for Lewis' rejection of the existence of genuine inconsistent worlds is the success of this *reductio ad absurdum*, then his argument fails, because the attempted *reductio ad absurdum* from assuming the existence of *w** is unsound (since P1 is false). Consequently, using only Lewis's reasons for rejecting the existence of genuine inconsistent worlds, it doesn't follow that they don't exist.

---

[137] The argument has the form of *reductio ad absurdum*.

[138] An analogue of this argument has been made in the context of extended theories of objects that include non-existent objects, e.g. Parsons's theory – a succinct summary of Parsons's analogue to the foregoing argument is given in (Zalta 1988, p.132).

[139] An analogous issue arises, concerning a questionable principle in theories of non-existent objects, e.g. Parsons (1980, p.19, p.105), Zalta (1988, pp.131-4). The analogy is along the lines *possible* vs. *impossible* and *existent* vs. *non-existent*. Parsons points out that the analogous principle to MCN holds for all *existing* objects, but not *all* objects (where impossible ones are among the non-existent). In that theory inconsistent objects are classified as non-existent objects, much in the same way as Lewis would classify them as classically impossible.

For similar reasons, a "Lewis-type" *reductio ad absurdum* against the existence of genuine *incomplete* worlds isn't sound either.[140] That is, given the breakdown of Lewis's argument above, the argument with the following conclusion is unsound, also due to P1' being false.

(Hyp') There is a world *w* such that it's neither the case that at *w*: *A* nor is it the case that at *w*: *not A*.
(P1') MCC and MCN are true of the domain restricting modifier for all worlds.
(C1') There is a world w such that at *w*: *A* and *not* at *w*: *A*. (Hyp+P1)
(P2') There is no subject matter whereby one can tell the truth by contradicting themselves (LNC).
(C2') Therefore, there is no such *w*.
(C3') Therefore there *are no* incomplete worlds. (RAA)
(C4') There are no impossible worlds.

Let us assume the existence of an incomplete world *w*° that has, as one of its parts, an object *a* such that neither *Pa* nor *not Pa* is true at *w*°, and then let us investigate what is *really* doing the work in a "successful" derivation of a contradiction. That is, *Pa* is a truth value gap at *w*°, so we have both '*not* at *w*° *Pa*' and '*not* at *w*° *not Pa*'. How is MCN justified here, and can it be used to derive a contradiction? In particular, can we use MCN to infer 'at *w*°: *not Pa*' from *not* 'at *w*°: *Pa*'? Call that particular MCN instance MCNI for *importing* negation, i.e. going from *not* 'at *w A*' to 'at *w not A*', as the dual move to MCNE from the previous, inconsistent world *w** example, which hinged on *exporting* negation, i.e. from 'at *w not A*' to *not* 'at *w A*'.

(MCNI) *Importing negation*: If 'not at *w*: *A*', then 'at *w*: *not A*'.
(MCNE) *Exporting negation*: If 'at *w*: *not A*', then 'not at *w*: *A*'.

That would suffice for a derivation of a contradiction, since we already have '*not* at *w*° *not Pa*'. This is also necessary since I have already argued that the other direction, i.e. MCNE is unwarranted, thereby disabling the only other way of deriving a contradiction. (In passing we can quickly set out that other means of deriving a contradiction, were MCNE to be allowed in the case of incomplete worlds. Given MCNE we can go from *not* 'at *w*° *not Pa*' to *not* [*not* 'at *w*° *Pa*'], and from there to 'at *w*° *Pa*', and given that we also have *not* 'at *w*° *Pa*', a contradiction ensues. But MCNE fails (as argued earlier), so let's focus on the inapplicability of MCNI in the case at hand.) The use of MCNI would seem wrong, since *w*° is an

[140] The next two paragraphs can be skipped.

incomplete world, so we cannot assume that the failure of *Pa* being true at *w*° means that *not Pa* should be true there. To do so would be to assume that there *is* a logical connection between the failure of *Pa* being true at *w*° and the truth of *not Pa* at *w*° after all. As a matter of fact, it would amount to assuming that LEM ought to hold at *w*°, but it *needn't* because *w*° isn't a classically possible world, and as a matter of fact it *doesn't*, precisely because *Pa* is a truth value gap at *w*°.

So, the above shows that Lewis-type arguments against the existence of impossible worlds fail, because they fail to rule out the existence of inconsistent and incomplete worlds, which are classically impossible.

3.5 No trouble: the 'marvellous mountain' argument is *invalid*.

The objection that shows 'the marvellous mountain' argument to be unsound can be modified to showing the argument to be invalid if we observe that assuming (and that what Lewis does) that MCC and MCN should always hold of the restricting modifier, amounts to assuming that there are no worlds where MCC or MCN fails to hold of the modifier, i.e. there are *no* inconsistent or incomplete worlds. Insisting on MCN always holding of the modifier, amounts to insisting that all worlds obey LNC, which consequently rules out there being any inconsistent worlds – the precise and particular conclusion that Lewis argues for (C3). But having already employed a principle (MCN) that doesn't hold for inconsistent worlds, *Lewis has effectively assumed that there are no such worlds* (since the use of MCN is blatantly illicit at such worlds), in particular the world from the hypothesis, where the marvellous mountain is said to exist. This begs the question against the existence of genuinely inconsistent individuals.

Consider the contrast in the applicability of MCN in the two following cases; the first based on an example from Lewis (Lewis 1986), where MCN intuitively and uncontroversially applies, and the second one borrowed from Priest's story Sylvan's Box (Priest 1997) where it just obviously cannot apply, since the box is stipulated to be an impossible object. That is, the *box from* the story is *both* empty *and* contains a wooden figurine fixed to its bottom, whereas the fridge is a classical object, as are its contents.

(1) In the fridge: *there is no beer*.

(2) In the box: *there is no wooden figurine*.

These two propositions have the negation within the scope of their respective domain restricting modifiers. Now if MCN was valid for both classical and non-classical domains its application would warrant *salva veritate* the following transformations of (1) and (2):[141]

(3) It's not the case that <u>in the fridge</u>: *there is some beer*.

(4) It's not the case that <u>in the box</u>: *there is a wooden figurine*.

It's rather obvious that the move from (1) to (3) via recourse to MCN is not only natural and intuitive, but justified, under the (stipulated) assumption that the fridge and its contents are classical – in particular they obey LNC. In fact the two propositions are classically equivalent – *(1) is true if and only if (3) is true*. By contrast, given that the box is what it is – an impossibilia of the *truth value glut* kind, (2) doesn't imply (4) *because whereas (2) is true (4) is as a matter of fact false (and false only), and its negation is true. Hence assuming MCN as valid rules out talk of inconsistent worlds, because we have just seen such worlds are counterexamples to MCN's validity. The fallacy is clear once we reveal all the enthymematic content that the premises carry:*

(Hyp) There is an inconsistent world w such that <u>at w</u>: *A* and *not A*.

(P1) MCC and MCN are true of the domain restricting modifier for all worlds, which implies that there are no inconsistent worlds (question begging).

(C1) There is a world w such that <u>at w</u>: *A* and *not* <u>at w</u>: *A*. (Hyp+P1)

(P2) There is no subject matter whereby one can tell the truth by contradicting themselves (LNC).

(C2) Therefore, there is no such w.

(C3) Therefore there *are no* inconsistent worlds. (RAA)

(C4) There are no impossible worlds.

3.6 Conclusion

In conclusion let's clearly emphasize what the above means for the extended argument from admissible paraphrase. Let's recall that the explicit position of Lewis (1973) is that such face-value interpretation ought to be regarded as affording the best semantic theory of the discourse if it does not lead to trouble and the alternatives do (Divers 2002, p.68):

[141] I'm aware of the implicit fictionalist account of possible and impossible worlds, which the story can be seen to endorse. The box from the story is to be considered as a Lewisean kind of *impossibilia*, i.e. real and concrete. I only use this example for historical reasons

> I do not make it an inviolable principle to take seeming existential quantifications in ordinary language at their face value. But I do recognize a presumption in favor of taking sentences at their face value, unless (1) taking them at face value is known to lead to trouble, and (2) taking them some other way is known not to. (Lewis 1973, p.84)

It has been shown that, on the extended account, the most obvious kind of trouble to Lewis—an inconsistent theory—does not arise, so although not entirely dismissed (1) is substantially weakened.

Chapter 4

Ordering Semantics for Counterfactuals
&
A Contextualized Account of Counterfactuals

> *We may separate the contribution of practice and context from the contribution of the world, evaluating counterfactuals as true or false at a world, and according to a 'frame' determined somehow by practice and context.*
>
> (Lewis, 1981)

> *Our system of spheres is nothing but a convenient device for carrying information about the comparative similarity of worlds. We could do away with the spheres, and give the truth conditions for counterfactuals directly in terms of comparative similarity of worlds* […]
>
> (Lewis 1973)

4.0 Introduction

This chapter gives an account of a contextualized (context relativized) counterfactual of the form 'In context C: *If it were the case that … , then it would be the case that …*', based on Lewis' (1974, 1981) analysis of the counterfactual. Drawing on earlier work by Lewis (1973, 1981) I first give an ordering semantics for counterfactuals, based on the idea of comparative similarity, interpreted as 'similarity in *relevant* respects' or as '*relevant* similarity', and modelled by total preorderings of possible worlds. Subsequently, building on that analysis I develop model-theoretic methods for a semantic consequence relation of contextualized counterfactuals (contextualized validity), which is given as the culminating item of the chapter.

The early part of the chapter (§4.1-§4.2) is devoted to defining the counterfactual language and an ordering semantics model theory, where comparative similarity of worlds is modelled by total preorders. In section §4.2, drawing on arguments given earlier in chapter 2, the resulting logic **CS** that I endorse is much like Lewis' preferred account of the counterfactual save for strict centering being replaced with a weaker centering condition. That is, **CS** is just the logic that Lewis (1973) calls **VW**, which is obtained from Lewis' preferred system **VC** (commonly referred to as **C1**) by replacing the *strict centering* condition with the *weak centering* condition, or equivalently removing the axiom $(A \wedge B) \supset (A > B)$ from the axiomatized version of **VC** (Lewis 1973, p.132). The system **CS** has a special relationship to the system **S** of chapter 2. The appendix on page 163 contains the proof of the equivalence of the class of **CS** models and the class of **S** models. That is, these classes validate the same sets of formulae and inferences.

In section §4.3 I further develop those features of **CS** model theory that are designed to provide a foundation for the logics of contextualized counterfactuals **CS**+, **CS1**+ and **CS2**+ developed in sections §4.4.3-§4.4.5. Because those features are central in their significance, they deserve a fitting prelude at this point.

Ordering frames, which constitute the basis of **CS** model theory given in §4.2 are – much like systems of spheres – a means of carrying information about the comparative similarity of worlds, relative to the actual world (or any other world where a counterfactual's truth is being evaluated). On Lewis' (1981, §2) conception of comparative similarity (which I adopt) as being largely determined by contextual considerations – contingent both on the facts that obtain at the actual world and what (counter-facts) we deem as relevant in any given conversational setting – ordering frames can be viewed as *carriers of contextual information*. On the reading of similarity as *similarity in relevant respects*, which I also maintain, comparative similarity of worlds is closely tied to relevance. Just as we think of possible worlds as the ways the world could be, we can think of ordering frames as ways that all facts and propositions could be distributed as a function of their relevance (importance, significance) to any given conversation.
This role of ordering frames – as contextual information carriers – can be developed further, in order to account for transformations (on ordering frames) interpreted as adding or subtracting contextual information. In §4.3 I define special classes of ordering frames, viz.

refinements and their duals, *dilutions*, which result from adding or subtracting contextual information, respectively to/from other ordering frames.
As an introductory example, consider the domain of integers and let us stipulate absolute value as the sole relevant parameter of "similarity to zero" – the smaller the absolute value, the greater the similarity to zero. Then any integer *n* and its negative counterpart *-n* are said to be *equivalent in terms of their comparative similarity to zero*, yielding the following ordering, (where the leftmost item indicates zero as being the most similar integer to itself, and rightward being the direction of increasing dissimilarity to zero):

(4.1) $\{0\}, \{-1, 1\}, \{-2, 2\}, \ldots, \{-n, n\}, \ldots$

This ordering (which is a *total preordering* of the integers) can be said to carry the information about all integers' relevant similarity to zero, where *relevant* similarity in this case is the integer's absolute value. We could *add more information* to this ordering – that is, we could *refine* the information in this ordering by adding a parameter that would introduce distinctions where none previously existed on account of their irrelevance, i.e. +/- sign distinction irrelevance.

A *refinement* of (4.1) would be an ordering that results from introducing new distinctions (interpreted as *adding contextual information* to the original ordering), whilst preserving the previous relevant comparative similarity differences (interpreted as *preserving contextual information* carried by the original ordering).[142] For example, consider a refinement in which all positive integers are taken to be more similar to zero than their negative counterparts:

(4.2) $\{0\}, \{1\}, \{-1\}, \{2\}, \{-2\}, \ldots, \{n\}, \{-n\}, \ldots$

Note that the original comparative similarity distinctions (original contextual information) have been preserved, i.e. as before any *m* is more similar to zero than any *n* iff $|m|<|n|$.
The central idea of refinements immediately gives rise to a more general and equally important concept of *mutual refinements* which are refinements of more than one ordering frame, and which are central to the notion of *contextual information preservation* underlying the semantic consequence relation defined on the contextualized language in §4.4.5.

[142] There are rudimentary parallels between what I call refinements and set partition refinements.

Sections §4.4.1-§4.4.3 constitute the model theory of the proposed analysis of contextualized counterfactuals, consisting of context representation, a formal language and its semantics. Setting up the basics of the semantics for the contextualized language in §4.4.1 I designate the role of context representation to **CS** ordering frames (which constitute the basis of the **CS** account of counterfactuals given earlier) and argue that they are adequate for that purpose.

The formal language for contextualized counterfactuals, given in section §4.4.2, introduces context-indexed connectives $>_c$ for each context *c*. That is, expressions like $A >_c B$ in the formal language intend to model contextualized counterfactuals of the form 'In context c: *If it were the case that A, then then it would be the case that B'*, where A and B express propositions. Subsequently, the corresponding semantics (**CS**+ model theory) of thus contextualized language allows making distinctions in the truth value of counterfactuals with the same antecedents (and even the same antecedents *and* consequents), by appeal to contextual considerations explicitly indicated by their respective context indices.

The semantics for the contextualized language, is developed in §4.4.2 and §4.4.3, and draws strongly on **CS** model theory (intended to serve as the foundation for **CS**+ model theory), i.e. by developing a mechanism that reduces the truth conditions of $A >_c B$ on a **CS**+ model to those of $A > B$ on a corresponding **CS** model whose underlying ordering frame is taken to represent context *c*.[143] That is, contextual considerations underlying a context-indexed expression are cashed out in terms of contextual information carried by ordering frames. The greater expressive power of the contextualized formal language gives the proposed analysis clear advantages over some alternative accounts.[144]

The culmination of the chapter is the logic of contextualized counterfactuals **CS2**+, offered in the form of a semantic consequence (contextualized validity), developed throughout sections §4.4.3-§4.4.5. I begin with the most basic system of the contextualized language **CS**+, defined at the end of f §4.3, which is hardly a logic of contextualized counterfactuals, since it

[143] The general character of the model theory of the contextualized language doesn't require the use any particular counterfactual analysis as its basis. The formalism is not entirely dependent on the base conditional logic. The proposal is a general prescription of how to contextualize a conditional language. The advantage of CS are the results about ordering frames and mutual refinements that allow to fashion a meaningful notion of contextual information preservation when defining semantic consequence.

[144] Gabbay's (1972) account has allowed for distinguishing between counterfactuals with same antecedents but different consequents, however falls short of distinguishing counterfactuals with the same antecedent and consequent, an observation also made by Nute (1980, p.76).

doesn't place any formal validity constraints on semantic consequence that require a contextual link between the premises and conclusion.

To develop such a notion of a contextual connection between the premises and the conclusion – a form of contextual information preservation – I show in §4.4.4 that **CS**+ preserves much of **CS**, if certain contextual restrictions are in place, and then I define the first proper logic of contextualized counterfactual **CS1**+ at the beginning of §4.4.5 with those contextual constraints included. To give a sense the contextual constraints placed on **CS**1+ semantic consequence, let us consider the following example:

In context *a* : If Caesar had been in command, he would have used the A-bomb.
In context *b* : If Caesar had been in command, he would have used catapults.
Therefore
In context *c* : If Caesar had been in command, he would have used the A-bomb and catapults.

The fundamental idea underlying contextualized validity is to add a condition of the existence of a link between the contextual information underlying the premises and the conclusion. To put it simply, for the contextualized inference to be valid it is required that the conclusion context *c* preserves the contextual information of the contexts *a* and *b* that make the premises true.[145] Only then a truth preserving inference can be said to be contextually valid. There is an interesting parallel between this requirement and the *syntactic* necessary condition for valid relevant conditionals that we have encountered earlier (chapter 1, *Definition 1.7*), which demands that the antecedent and consequent share a common propositional variable. Here, on the other hand, we have an analogous *semantic* necessary condition for valid inference, which demands that the premises and the conclusion share a common structure, interpreted as a carrier of contextual information.

[145] Berto (2014, p.113; 2017, p.11) notices – in the context of a logic of imagination – that the conclusion shouldn't come automatically, as a logical entailment. However, the solution he suggests (Berto 2014, p.113, f.9; Berto 2017, p.11), of fixing the premises to range across a single context, although sound, needn't be that strong. The challenge, as I see it is to allow premises to range over more than one context and propose a means for contextual information that ensures truth preservation. There are some parallels between my conditions for *contextualized validity* and those suggested by Priest (2017, §3.2), for what he calls *material validity*, albeit expressed in terms of *imported information*.

I conclude §4.4.5 and the chapter by defining the system **CS2**+, which is weaker than **CS1**+ due to an alteration of the contextual constraints – one motivated by the invalidation of some questionable inferences whose formal validity **CS1**+ inherits from **CS**.

The main results of section §4.3, devoted to the model theory based on ordering frames, are *Lemma 4.3*, regarding the duality between refinements and dilutions, *Proposition 4.5* which establishes important truth preserving property of refinements, and the dual result concerning falsity preserving properties of dilutions is given by *Corollary 4.5.1*. The main results in section §4.4 of the modified model theory for the contextualized language are *Theorem 4.9*, and *Theorem 4.10*, which say that depending on the extent of restrictions on the kind of context indices appearing in the premises and conclusion, **CS**+ preserves either all of the logic **CS**, or some of it. *Corollaries 4.9.1* and *4.9.2* say that **CS**+ reduces to **S5** or **CS** if the context index-set is empty or a singleton, respectively.

4.1 The formal language

First let's start with the basic ingredients for our language, i.e. a set of propositional variables $PV = \{p_n : n \in \mathbb{N}\}$ the elements of which shall be denoted with lowercase Roman letters $(p, q, r, \dots)$ or subscripted lowercase Roman p's $(p_1, p_2, \dots, p_k, \dots)$, or lowercase Greek letters $(\varphi, \psi, \chi, \dots)$; unary connectives: $\sim$ (negation), $\Box$ (necessity), $\Diamond$ (possibility); and binary connectives: $\wedge$ (conjunction), $\vee$ (disjunction), $\supset$ (material conditional), $>$ (counterfactual conditional). For the metalanguage, upper case letters $(A, B, C, \dots)$ shall be used as variables ranging over complex formulae and propositional variables.

Definition 4.1.1: Define our language of interest, denoted $\mathcal{L}$, to be the set: $\{\sim, \Box, \Diamond, \wedge, \vee, \supset, >\}$.

Definition 4.1.2: Let the set of propositional variables be $PV = \{p_n : n \in \mathbb{N}\}$.

Now we define the set of well-formed formulae.[146]

Definition 4.1.3: Let For be the smallest set closed under the following well-formed formula formation rules:

[146] E.g. the counterfactual 'If kangaroos had no tails, they would topple over' would have the form: $p > q$, where p stands for 'kangaroos have no tails' and q stands for 'kangaroos topple over'. This is the same language, as defined in §2.1.1.

B: All propositional variables are wffs, i.e. $PV \subseteq For$.

R1: If $A \in For$ then $\{\sim A, \Box A, \Diamond A\} \subseteq For$.

R2: If $\{A, B\} \subseteq For$ then $\{A \wedge B, A \vee B, A \supset B, A > B\} \subseteq For$.

Definition 4.1.4: $A \in For$. The *set of subformulae of A* is the smallest set $Sub(A)$ satisfying the following conditions:

1. $A \in Sub(A)$
2. For each $* \in \{\sim, \Box, \Diamond\}$ if $* B \in Sub(A)$, then $B \in Sub(A)$.
3. For each $\circ \in \{\wedge, \vee, \supset, >\}$ if $B \circ C \in Sub(A)$, then $B \in Sub(A)$ and $C \in Sub(A)$.

Next, we proceed to defining structures on which $>$ is to be defined. The conditions that have been adopted are those thought to be best in terms of what Lewis' account offers, i.e. what inferences it validates and which ones it invalidates, as discussed in *Chapter 2*.

Note that for all $A, B \in For$: if $B \in Sub(A)$, then $Sub(B) \subseteq Sub(A)$.
Proof : It's immediate, from the definition.

Definition 4.1.5: It will be helpful to define the subset of For that contains all and only formulae that contain occurrences of $>$. Denote that subset with $For_>$.
More formally: $A \in For_>$ iff $\exists B \in Sub(A)$ such that $B = C > D$ for some $C, D \in For$.

Definition 4.1.6: Denote the set $For \backslash For_>$ with $\overline{For_>}$, which is just the set of wffs of basic modal language.

Definition 4.1.7: Define $For_{>0} \subseteq For_>$ as follows: $C \in For_{>0}$ iff whenever $A > B \in Sub(C)$, then both $A \in \overline{For_>}$ and $B \in \overline{For_>}$.

That is, $For_{>0}$ is just like $For_>$, but any instances of $A > B$ are restricted in the above sense.
Example: $\sim(p > (q \supset r)) \wedge \Big(((p \wedge \sim q) > r) \vee (q > \sim r)\Big) \in For_{>0}$ but $p > (p > p) \notin For_{>0}$.

Definition 4.1.8: Define $For(>) \coloneqq \{A > B : A, B \in For\}$. That is, $For(>)$ is just the set of For formulae whose main connective is $>$.

4.2 Comparative Similarity

In order to know what the relations in our semantics are, we need to introduce their intended meaning and basic properties. The systems of spheres discussed in §2.2 are just a convenient, and intuitive way for representing information about the comparative similarity of worlds.[147] We can do the same, directly in terms of comparative similarity of worlds, together with accessibility. To make this explicit let's consider the following definitions.

Definition 4.2.1: A binary relation $R \subseteq S \times S$ on a set S, denoted by $\lesssim$, is a *preorder* iff it is:

(1) *Transitive*: $\forall x, y, z \in S \left((x \lesssim y \wedge y \lesssim z) \rightarrow x \lesssim z\right)$.

(2) *Reflexive*: $\forall x \in S\ (x \lesssim x)$.

If $\lesssim$ satisfies (1), (2), and (3), it is a *total preorder* (also called a *non-strict weak order*).

(3) *Totality*: $\forall x, y \in S\ (x \lesssim y \vee y \lesssim x)$. [148]

Definition 4.2.1.1: For any preorder $\lesssim$, denote $(x, y) \notin \lesssim$, i.e. 'it is not the case that $x \lesssim y$' with $y < x$, and let us write $x \sim y$ to mean that both $x \lesssim y$ and $y \lesssim x$.

Corollary 4.0.1: If $\lesssim$ is a *preorder* on S then for no $x \in S$: $x < x$.
Proof : This follows directly from reflexivity of $\lesssim$, i.e. $x < x$ means $(x, x) \notin \lesssim$, contradicting reflexivity of $\lesssim$. □

Corollary 4.0.2: If $\lesssim$ is a *total preorder* on S then for all $y, x \in S$:

(i) $x < y$ iff $(x, y) \in \lesssim$ and $(y, x) \notin \lesssim$

(ii) $x \lesssim y$ iff $x < y$ or $x \sim y$

Proof : (i) $(y, x) \notin \lesssim$ follows from definition of $x < y$, and $(x, y) \in \lesssim$ follows from totality of $\lesssim$. (ii) Given totality, either $(x, y) \in \lesssim$ and $(y, x) \notin \lesssim$ or both $(x, y) \in \lesssim$ and $(y, x) \in \lesssim$. The third, totality satisfying option $(x, y) \notin \lesssim$ and $(y, x) \in \lesssim$ is clearly impossible. □

My definition of ordering frames based on comparative similarity closely follows the definition of *comparative similarity system* in Lewis (1973, p.48), save for the condition corresponding to what Lewis calls *centering*, i.e.

---

[147] Lewis (1973, p.48).
[148] Lewis (1973, p.48) refers to this property as *strongly connected*.

(CS3.1) The element i is $<_i$-minimal: $\forall j \in W(j \neq i \longrightarrow i <_i j)$.

which I replace with a weaker condition (CS3) corresponding to *weak centering* for reasons already given in §2.2.5.

Definition 4.2.2: An *ordering frame* based on comparative similarity is a pair $(W, \lesssim)$, where W is a nonempty set and $\lesssim: W \longrightarrow \wp(W) \times \wp(W \times W)$ is a function that assigns to each $i \in W$ a pair $(S_i, \lesssim_i)$, consisting of a set $S_i \subseteq W$, regarded as the set of worlds accessible from i, and a binary relation $\lesssim_i$ on W, regarded as the ordering of worlds in respect of their comparative similarity to i and satisfying the following conditions, for each $i \in W$:

(CS1) $\lesssim_i$ is a total preorder on W

(CS2) i is self-accessible: $i \in S_i$.

(CS3) i is $\lesssim_i$-minimal: $\forall j \in W(i \lesssim_i j)$.

(CS4) Inaccessible worlds are $\lesssim_i$-maximal: $\forall j, k \in W(k \notin S_i \longrightarrow j \lesssim_i k)$.

(CS5) Accessible worlds are more similar to i than inaccessible worlds:

$$\forall j, k \in W\big((j \in S_i \wedge k \notin S_i) \longrightarrow j <_i k\big)$$

On the intended interpretation, elements of W are possible worlds, S_i is regarded as the set of worlds accessible from i, and $\lesssim_i$ is regarded as the ordering of worlds in respect of their comparative similarity to i, with the following intended meaning:

$j \lesssim_i k$: j is at least as similar to i as k is.

$j <_i k$: j is more similar to i than k is.

$j \sim_i k$: j and k are equally similar to i.[149]

Definition 4.2.3: Denote the *class of ordering frames from Definition 4.2.2* by **CS**.
Note that since *centering* implies *weak centering*, the class of ordering frames where we substitute (CS3) for *centering* is a proper subclass of **CS**.[150]

[149] Lewis' (1981, p.220) definition of $\sim_i$ in terms of a strict comparative similarity relation $<_i$ is logically equivalent to the one he gave earlier, in Lewis (1973, p.48) – the one I choose to use in the remainder of the chapter. In terms of $<_i$ the comparative similarity equivalence $\sim_i$ is defined as follows: $j \sim_i k$: neither $j <_i k$ nor $j <_i k$.

[150] Since, if $j <_i^F k$, then $j \lesssim_i^F k$ for any $i, j, k \in W$, by totality and definition of $<_i^F$.

Definition 4.2.4: Given some $F \in$ **CS**, let W^F denote the domain of F and let $\lesssim^F$ denote F's ordering/accessibility assignment on F's domain, i.e. $W^F \longrightarrow \wp(W^F) \times \wp(W^F \times W^F)$ as defined in 4.2.2. Also, let S_i^F and $\lesssim_i^F$ denote the elements of the image $\left(S_i^F, \lesssim_i^F\right)$ of $i \in W^F$ under $\lesssim^F$.

It may be of use to define comparative similarity equivalence classes.

Definition 4.2.4.1: Let $[\![k]\!]_i^F = \left\{j \sim_i^F k : j \in S_i^F\right\}$ for $F \in$ **CS** and all $i \in W^F$.
That is, $[\![k]\!]_i^F$ is the similarity equivalence class of worlds that are equally similar to i as k is, according to the ordering assignment $\left(S_i^F, \lesssim_i^F\right)$.[151]

To define the notion of truth according to an ordering frame, we need to define models. I'm tempted to adopt a valuation *relation* from Dunn's (1976) semantics for FDE, which will facilitate generalizing the model theory once impossible worlds are introduced in *Chapter 5*. That is, use a valuation relation $\rho \subseteq (W \times PV) \times \{0,1\}$, and for Lewis' account restrict ρ so it is a function.[152]

Definition 4.2.5: A model *based on comparative similarity* is the triple $(W, \lesssim, \rho)$ such that $(W, \lesssim)$ is an *ordering frame* and for each $i \in W$, $\rho_i \subseteq PV \times \{0,1\}$ is a relation between PV and $\{0,1\}$. Informally we think of $\{i \in W : p\rho_i 1\}$ as the set of worlds in the model where is p true, and $\{i \in W : p\rho_i 0\}$ as the set of worlds in the model where is p false.
For the duration of this chapter, we constrain ρ so that for all $p \in PV$ and $i \in W$ either $p\rho_i 0$ or $p\rho_i 1$, but not both. That is, for the time being we place the following constraints on ρ:

Exclusion: for no $p \in PV$ and $i \in W$, both $p\rho_i 0$ and $p\rho_i 1$.
Exhaustion: for all $p \in PV$ and $i \in W$, either $p\rho_i 0$ or $p\rho_i 1$.[153]

Note that thus restricted ρ is effectively a function $\rho_i : PV \longrightarrow \{0,1\}$ for each $i \in W$. Truth in a model is defined in terms the satisfiability relation $\Vdash \subseteq W \times For$. We read $i \Vdash A$ as 'A is true at i'. Given a model $(W, \lesssim, \rho)$ and any $i \in W$, define $\Vdash$ as follows:

(1) $i \Vdash p$ iff $p\rho_i 1$

[151] Note that for every assignment $(S_i^F, \lesssim_i^F)$, the set of all $\sim_i^F$ equivalence classes $\{[\![j]\!]_i^F : j \in S_i^F\}$ partitions S_i^F.
[152] To be precise, I will use Priest's reconstruction of Dunn's (1976) formulation, where in the latter truth values are identified with subsets of $\{0,1\}$.
[153] Borrowed from Priest (2008, §8.4).

(2) $i \Vdash \sim A$ iff not $i \Vdash A$

(3) $i \Vdash A \wedge B$ iff $i \Vdash A$ and $i \Vdash B$

(4) $i \Vdash A \vee B$ iff $i \Vdash A$ or $i \Vdash B$

(5) $i \Vdash A \supset B$ iff $i \Vdash \sim A$ or $i \Vdash B$

(6) $i \Vdash \Box A$ iff $\forall j \in W\colon j \Vdash A$.

(7) $i \Vdash \Diamond A$ iff $\exists j \in W\colon j \Vdash A$.

(8) $i \Vdash A > B$ iff (i) $\sim\exists k \in S_i\colon k \Vdash A$, or

(ii) $\exists k \in S_i\colon k \Vdash A$ and $\forall j \in S_i\,(j \lesssim_i k \longrightarrow j \Vdash A \supset B)$

- Any reference to ρ for the remainder of this chapter assumes ρ satisfying both *exclusion* and *exhaustion*. For convenience, let's introduce the following notation:

  $i \Vdash \Sigma$ iff $i \Vdash A$ for all $A \in \Sigma$

- When we want to explicitly refer to truth at a world in a particular model $\mathfrak{A}$, we shall employ the following notation: $\mathfrak{A}, i \Vdash A$ and $\mathfrak{A}, i \Vdash \Sigma$.

- Also denote with $\mathfrak{A} \Vdash A$ when $\mathfrak{A}, i \Vdash A$ for all $i \in W^{\mathfrak{A}}$.

Definition 4.2.6: It will also be convenient to define $[A]^{\mathfrak{A}} := \{i \in W\colon \mathfrak{A}, i \Vdash A\}$ for any model $\mathfrak{A}$ with domain W. The superscript will be omitted in cases when its absence will not lead to ambiguity.

Definition 4.2.7: Let $\vDash_{\mathbf{CS}} \subseteq \wp(For) \times For$, and define $\Sigma \vDash_{\mathbf{CS}} A$ iff for all models $(W, \lesssim, \rho)$, and all $i \in W$, if $i \Vdash B$ for all $B \in \Sigma$, then $i \Vdash A$. We say an inference from Σ to A is valid iff $\Sigma \vDash_{\mathbf{CS}} A$. That is, valid inference is defined as truth preservation at all worlds in all **CS**-models. A formula $A \in For$ is said to be valid iff $\emptyset \vDash_{\mathbf{CS}} A$. Call this logic **CS**.

Note that since the truth conditions for $\Box$ and $\Diamond$ formulae are defined in terms of unrestricted quantification over possible worlds, i.e. only $>$-formulae truth conditions contain accessibility restrictions, the above validity conditions give the modal logic **S5** for the basic modal language.

Just as we have relativized formula validity to a model $\mathfrak{A} \Vdash A$ it will be of use to define valid inference relativized to a model.

Definition 4.2.7.1: Let $\vDash_{\mathfrak{A}} \subseteq \wp(For) \times For$, and given a **CS** model $\mathfrak{A} = (W, \lesssim, \rho)$ write

- $\vDash_{\mathfrak{A}} A$ iff $\mathfrak{A} \Vdash A$
- $\Sigma \vDash_{\mathfrak{A}} A$ iff for all $i \in W$, if $\mathfrak{A}, i \Vdash B$ for all $B \in \Sigma$, then $\mathfrak{A}, i \Vdash A$.

This allows us to give a more succinct definition of semantic consequence:

$$\Sigma \vDash_{\mathbf{CS}} A \quad \text{iff} \quad \text{for all } \mathbf{CS} \text{ models } \mathfrak{A}\text{: } \Sigma \vDash_{\mathfrak{A}} A$$

Note that it is immediate from the above definitions that $\vDash_{\mathbf{CS}} \subseteq \vDash_{\mathfrak{A}}$.

4.3 Ordering frame refinements and dilutions

Let us start now turn to defining ordering frame refinements and dilutions.[154]

Definition 4.3.1: Let $\mathcal{R} \subseteq \mathbf{CS} \times \mathbf{CS}$ and call an ordering frame G a *refinement* of ordering frame F iff $(F, G) \in \mathcal{R}$. And define $(F, G) \in \mathcal{R}$ iff:

(i) $W^G = W^F$,

and for all $i \in W^F$:

(ii) $\lesssim_i^G \subseteq \lesssim_i^F$

(iii) $S_i^G = S_i^F$

Definition 4.3.1.1: A *proper* refinement of F is a refinement G, such that $G \neq F$.

Definition 4.3.1.2: Let $\mathcal{R}[F] \coloneqq \{G \in \mathbf{CS}: (F, G) \in \mathcal{R}\}$ denote the *image* of F under $\mathcal{R}$, i.e. the set of all refinements of F.

Definition 4.3.2: Let $\mathcal{D} \subseteq \mathbf{CS} \times \mathbf{CS}$ and call an ordering frame G a *dilution* of ordering frame F iff $(F, G) \in \mathcal{D}$. And define $(F, G) \in \mathcal{D}$ iff:

(i) $W^G = W^F$,

and for all $i \in W^F$:

(ii) $\lesssim_i^F \subseteq \lesssim_i^G$

(iii) $S_i^G = S_i^F$

[154] The essential idea of refinements is based on Lewis (1981, pp.226-7). However, Lewis (1981) defines refinements on strict preorder relations: if $j <_i^F k$, then $j <_i^G k$ (where G is a refinement of F). Given the way I have defined refinements (using total preorders) Lewis' definition is a derived property of refinements, i.e. *Lemma 4.1*.

Definition 4.3.2.1: A *proper* dilution of F is a dilution G of F, such that $G \neq F$.

NOTE: the orderings of refinements and dilutions are total, by definition of ordering frames.

Definition 4.3.2.2: Let $\mathcal{D}[F] \coloneqq \{G \in \mathbf{CS}: (F, G) \in \mathcal{D}\}$ denote the *image* of F under $\mathcal{D}$, i.e. the set of all dilutions of F.

4.3.1 Intended role and meaning of ordering frame refinements & dilutions

The following informal discussion intends to give a better understanding of what refinements and dilutions are, and how it is precisely that they represent what they are intended to represent. It includes a number of examples and important limit cases. It is important that the readers' intuitions about refinements and dilutions are secured, since these structures are central to most of the content in this chapter.

4.3.1.1 Representing total preorders

If will be useful to employ useful representations of total preorders in our discussion to aid the explanation how they are intended to carry contextual information. The definition below offers an intuitive means of representing total preorders.

Definition 4.3.3: Let $(S, \lesssim)$ be a total preorder. Let $a \mid b$ represent $a < b$ and $a\ b$ or $b\ a$ (or with commas, e.g. a, b or a, b) represent $a \sim b$ for any $a, b \in S$, such that $a \neq b$.[155] Reflexivity in this picture is left as implicit. For larger, or infinite collections, I may write $a_1 \ldots a_n \mid b_1 b_2 \ldots$

Example: Let $(W, \lesssim) = (\{a, b, c, d\}, \{(a, b), (a, c), (a, d), (b, c), (b, d), (c, b), (c, d), \ldots\})$, where the ellipsis denotes the remaining reflexive pairs. We say that $a \mid b\ c \mid d$ represents $(W, \lesssim)$. Note that $a \mid c\ b \mid d$ is the other valid representation.

Example: Let $P = (\mathbb{Z}, \lesssim_P)$ be the total preorder on the integers, defined: $m \lesssim_P n$ iff $|m| \leq |n|$. Intuitively we can say $m \lesssim n$ iff m is at least as close to zero as n. We can represent $(\mathbb{Z}, \lesssim)$ as:

$$0 \mid -1, 1 \mid -2, 2 \mid -3, 3 \mid \ldots$$

Clearly, $-n \sim_P n$ for all $n \in \mathbb{Z}$, and $m <_P n$ for all $m, n \in \mathbb{Z}$ such that $|m| < |n|$.

[155] Recall from definition 4.2.1.1 that for any preorder $\lesssim$ on a set S, we denote $(x, y) \notin \lesssim$, i.e. 'it is not the case that $x \lesssim y$' with $y < x$, and we write $x \sim y$ to mean that both $x \lesssim y$ and $y \lesssim x$.

4.3.1.2 Refinements and Dilutions

Strictly speaking, the terms *refinement* and *dilution* apply to *entire* ordering frames (which are more like families of total preorders), not the particular ordering assignments (similarity assignments). That is, ordering frames have the general structure $(W, \{(S_i, \lesssim_i) : i \in W\})$, and refinements are defined on such structures, but in the following explanation of the basic properties of refinements and dilutions I'll also extend the use of the term refinement/dilution, to the individual assignments themselves, i.e. if G is a refinement of F, I'll also refer to $(S_i^G, \lesssim_i^G)$ as the refinement of $(S_i^F, \lesssim_i^F)$ for some $i \in W$, since after all it is the ordering relationships of such individual assignments between ordering frames and their refinements/dilutions that are key. The following examples focus on the motivation and intended meaning of conditions (ii) of definitions 4.3.1 and 4.3.2 regarding the relationship of the particular ordering assignments to their refined/diluted counterparts. That is, we're going to discuss the meaning (formal and the intended interpretation) of the relationship between an assignment $\lesssim_i^F$ and its counterparts when the frame F is refined or diluted.

Example: Form the earlier example featuring $(\mathbb{Z}, \lesssim_P)$, there are infinitely many proper refinements, each resolving some tie in "closeness to zero", e.g.

$$0 \mid -1 \mid 1 \mid -2, 2 \mid -3, 3 \mid \ldots$$

or

$$0 \mid -1, 1 \mid 2 \mid -2 \mid -3, 3 \mid \ldots$$

or

$$0 \mid -1, 1 \mid \ldots \mid -k \mid k \mid \ldots$$

for some integer k. We could define a refinement where positive integers are deemed as more similar to zero than negative integers, despite their absolute value tie in closeness:

$$0 \mid 1 \mid -1 \mid 2 \mid -2 \mid 3 \mid -3 \mid \ldots$$

Or vice-versa:

$$0 \mid -1 \mid 1 \mid -2 \mid 2 \mid -3 \mid 3 \mid \ldots$$

Example: The total preorder on some world major cities is defined in terms of population size as compared to Brisbane (i.e. weakly centered on Brisbane), and we'll consider a refinement that includes distance to Brisbane as an additional similarity parameter. Below are the relevant similarity parameters, i.e. the demographics and distance from Brisbane (~ denotes *approx.*).

	Population (million):	Distance from Brisbane (1000 km):
Brisbane:	~2	0
Perth:	~2	~4.3
Auckland:	~2	~2.3
Warsaw:	~2	~15
Gold Coast:	~0.6	~0.1
Sydney:	~5	~1
Athens	~0.7	~15
St. Petersburg:	~5	~14

The similarity assignment **F**, or more precisely $(\{\mathrm{Br}, \mathrm{Per}, \mathrm{Auc}, \mathrm{War}, \mathrm{G.\,Coa}, \mathrm{Ath}, \mathrm{Syd}, \mathrm{St.\,Pet}\}, \lesssim^{F}_{\mathrm{Br}})$ carries the information about the comparative similarity of cities, relative to Brisbane, where the only relevant similarity parameter is a city's population.

The refinement **R1** of **F** takes distance as an additional relevant parameter, but in a very coarse-grained manner – distance differences relative to Brisbane within two thousand kilometres don't register as sufficiently relevant for the distinction i.e. Perth and Auckland aren't distinguished nor are Sydney and the Gold Coast, nor St. Petersburg and Athens.

The refinement **R2** of **R1** and *a fortiori* a refinement of **F**, introduces more resolution, making differences in distance relative to Brisbane below a thousand kilometres relevant, thereby distinguishing even the distance-to-Brisbane difference between Sydney and the Gold Coast, which is approximately a thousand kilometres.

F Brisbane, Perth, Auckland Warsaw | Gold-Coast, Athens, Sydney, St. Petersburg

R1 Brisbane, Perth, Auckland | Warsaw | Gold-Coast, Sydney | Athens, St. Petersburg

R2 Brisbane | Perth, Auckland | Warsaw | Gold-Coast | Sydney | Athens | St. Petersburg

Basically, in *total preorders* a proper refinement resolves at least a single symmetric pair, i.e. contains exactly *one* pair from the *two* contained in the original preorder, so for any total preorder $(W, \lesssim) = (W, \{(a, b), (b, a), \dots\})$, there exist refinements $(W, \lesssim^{\alpha}) = (W, \lesssim \setminus \{(a, b)\})$ and $(W, \lesssim^{\beta}) = (W, \lesssim \setminus \{(b, a)\})$. We interpret this as refinements *resolving* comparative similarity ties (symmetric pairs). Clearly, both $\lesssim^{\alpha}$ and $\lesssim^{\beta}$ are subsets of $\lesssim$. It follows that maximal (in the sense of most symmetric pairs being resolved) refinements are linear (in the sense that if $(W, \lesssim^{\alpha})$ is some maximal refinement of $(W, \lesssim)$, then $(a, b) \in \lesssim^{\alpha}$ and $(b, a) \in \lesssim^{\alpha}$

implies $a = b$). There are two important limit cases. Namely, each maximal refinement is a linear order on W, and there's a single maximal dilution i.e. $W \times W$.

Example: Let $(W, \lesssim) = (\{a, b, c, d\}, \{(a, b), (a, c), (a, d), (b, c), (b, d), (c, b), (c, d), \ldots\})$, where the ellipsis denotes the remaining reflexive pairs. Let's consider the only two proper refinements of $(W, \lesssim)$, namely $(W, \lesssim \setminus \{(c, b)\})$ and $(W, \lesssim \setminus \{(b, c)\})$.

$a \mid b\ c \mid d$	$(W, \lesssim)$
$a \mid b \mid c \mid d$	$(W, \lesssim \setminus \{(c, b)\})$
$a \mid c \mid b \mid d$	$(W, \lesssim \setminus \{(b, c)\})$

Clearly $\lesssim \setminus \{(c, b)\} \subseteq \lesssim$ and $\lesssim \setminus \{(b, c)\} \subseteq \lesssim$. Note that both refinements happen to be maximal refinements of $(W, \lesssim)$. Also, $(W, \lesssim)$ is a dilution of $(W, \lesssim \setminus \{(c, b)\})$ and $(W, \lesssim \setminus \{(b, c)\})$, by definition of dilutions.

4.3.1.3 Interpretation: Contextual Information

Refinements, whilst containing more contextual information (when we refine, we add contextual information by making additional distinctions), preserve the contextual information of the original ordering frame. Another way of looking at this is to view those distinctions (absent from the original ordering frame) as becoming relevant on the context represented by the refinement. Dilutions do the opposite – they remove previously existing distinctions, so when we dilute we are removing contextual information (irrelevant information), i.e. distinctions that have been relevant on the original frame are no longer relevant on the dilution.

Usually we tend to think of submodels as providing less information than their extensions. But in this case, there is a sense in which the opposite seems to be happening. When we refine, we are taking submodels, and we can keep going until we get to a linear ordering: that direction feels like we are adding information. On the other hand, if we take supermodels (dilute), the limit is the case where everything is related to everything else, which feels like we are losing information. This tends to go against the usual intuitions.[156]

[156] I owe this observation to Toby Meadows.

4.3.2 Properties of ordering frame refinements and dilutions

Now we resume the formal discussion and prove some basic properties of refinements and dilutions. Frame refinements preserve the strict ordering of original ordering frames in the following sense:

Lemma 4.1: If G is a refinement of F, then if $j <_i^F k$ for any i, j, k according to some comparative similarity assignment $(S_i^F, \lesssim_i^F)$, then $j <_i^G k$ according to $(S_i^G, \lesssim_i^G)$.

Proof : It suffices to note that, since $\lesssim_i^F$ is total and $\lesssim_i^G \subseteq \lesssim_i^F$ for each i, then if $(j,k) \in \lesssim_i^F$ and $(k,j) \notin \lesssim_i^F$, i.e. $j <_i^F k$, then it follows that both $(j,k) \in \lesssim_i^G$ and $(k,j) \notin \lesssim_i^G$, i.e. $j <_i^G k$. Denying $(k,j) \notin \lesssim_i^G$ contradicts the subset property, and $(j,k) \in \lesssim_i^G$ contradicts totality. □

We have a dual result to *Lemma 4.1* for frame dilutions. That is, frame dilutions preserve the non-strict ordering of original ordering frames in the following sense:

Lemma 4.2: If G is a dilution of F then if $j \lesssim_i^F k$ for any i, j, k according to some comparative similarity assignment $(S_i^F, \lesssim_i^F)$, then $j \lesssim_i^G k$ according to $(S_i^G, \lesssim_i^G)$.

Proof : It suffices to observe that, since $\lesssim_i^F \subseteq \lesssim_i^G$ for each i, if $(j,k) \in \lesssim_i^F$ then $(j,k) \in \lesssim_i^G$. □

Corollary 4.2.1: If $j \sim_i^F k$ for any i, j, k according to some comparative similarity assignment $(S_i^F, \lesssim_i^F)$ on a frame F, then $j \sim_i^G k$ according to any dilution G of F.

Proof : Immediate from *Lemma 4.2* and definition of $\sim_i$. □

The dual relationship between frame refinements and frame dilutions, although implicit in the definition, deserves highlighting.

Lemma 4.3: For any ordering frames $F, G \in \mathbf{CS}$, $(F,G) \in \mathcal{R}$ iff $(G,F) \in \mathcal{D}$.

Proof : It's immediate from definitions of refinements and dilutions. □

Lemma 4.4: For any ordering frames $F = (W^F, \lesssim^F)$, $G = (W^G, \lesssim^G)$, and any ρ:

If $W^F = W^G$ and $A \in \overline{For_>}$, then $(F, \rho), i \Vdash A$ iff $(G, \rho), i \Vdash A$.

Proof : It suffices to observe that the truth of formulae in $\overline{For_>}$ is independent of $\lesssim$. □

The result below is *central* to some key applications in this chapter. Refinements are truth-preserving in the following sense:[157]

Proposition 4.5: If a counterfactual $A > B \in For_{>0}$ is *true* at a world according to some ordering frame F, then it is true at that world according to any refinement of F. That is, for all $F = (W^F, \lesssim^F) \in \mathbf{CS}$, and for all $A, B \in \overline{For_>}$, $i \in W^F$, and ρ:

$$(F, \rho), i \Vdash A > B \quad \text{iff} \quad (\forall G \in \mathcal{R}[F])\big((G, \rho), i \Vdash A > B\big)$$

Proof : (←) Is immediate, since $F \in \mathcal{R}[F]$. (→) Consider some $F \in \mathbf{CS}$, $A \in \overline{For_>}$, $i \in W^F$, and ρ, such that $(F, \rho), i \Vdash A > B$, Hence, for all $A, B \in \overline{For_>}$, $i \in W^F$, ρ either $\sim\exists k \in S_i^F$: $(F, \rho), k \Vdash A$ or $\exists k \in S_i^F$: $(F, \rho), k \Vdash A$ and $\forall j \in S_i^F\big(j \lesssim_i^F k \longrightarrow (F, \rho), j \Vdash A \supset B\big)$. Let us start with the vacuous case (first disjunct) and assume for arbitrary $A \in \overline{For_>}$, $i \in W^F$, and ρ that $\sim\exists k \in S_i^F$: $(F, \rho), k \Vdash A$. From this, *Lemma 4.4*, and the fact that $S_i^G = S_i^F$ we can infer that $\sim\exists k \in S_i^G$: $(G, \rho), k \Vdash A$. Next, let us assume (the main hypothesis) $\exists k \in S_i^F$: $(F, \rho), k \Vdash A$ and $\forall j \in S_i^F\big(j \lesssim_i^F k \longrightarrow (F, \rho), j \Vdash A \supset B\big)$. To distinguish it from other assumptions call this assumption *the main hypothesis*. It follows that $\exists k \in S_i^G$ and $(G, \rho), k \Vdash A$ for all $G \in \mathcal{R}[F]$, by *Lemma 4.4* and the fact that $S_i^G = S_i^F$. Now, to show that $\forall j \in S_i^G\big(j \lesssim_i^G k \longrightarrow (G, \rho), j \Vdash A \supset B\big)$ we'll proceed by assuming $j \lesssim_i^G k$ for arbitrary $j \in S_i^G$, $G \in \mathcal{R}[F]$, and show $(G, \rho), j \Vdash A \supset B$. So, let's assume $j \lesssim_i^G k$ for arbitrary $j \in S_i^G$, $G \in \mathcal{R}[F]$, and note that since G is a refinement of F, then F is a dilution of G, by *Lemma 4.3*. Also, it should be noted that dilutions are $\lesssim$-preserving in the sense of *Lemma 4.2*. Hence, we conclude $j \lesssim_i^F k$, by *Lemma 4.2* and *4.3* . From this, and the main hypothesis we infer $(F, \rho), j \Vdash A \supset B$, which in conjunction with the fact that $W^F = W^G$ gives $(G, \rho), j \Vdash A \supset B$, by *Lemma 4.4*. Therefore, we finally conclude that $\forall j \in S_i^G\big(j \lesssim_i^G k \longrightarrow (G, \rho), j \Vdash A \supset B\big)$, by conditional proof.

This completes the proof. □

[157] Lewis (1981, pp.226-227) has proven a similar result. His result is more general than *Proposition 4.5* in one sense, and less general in another. Whereas *Proposition 4.5* holds only for a class of frames based on *total preorderings*, Lewis has proven a similar result for ordering frames based on *partial orderings* (where only refinements are required to be based on total preorderings). On the other hand, whereas Lewis has proven this only for (strongly) *centered* ordering frames, *Proposition 4.5* holds for *weakly centered* orderings frames, i.e. satisfying (CS3), so *a fortiori* it holds for ordering frames satisfying the (stronger) *centering* restriction (CS3.1). Also, the employment of frame dilutions and *Lemmas 4.2* and *4.3*. makes the proof of *Proposition 4.5* substantially simper than Lewis' proof.

We have a dual result for dilutions, which are falsity-preserving in the following sense:

Corollary 4.5.1: For all frames $F, G \in \mathbf{CS}$ and for all $A, B \in \overline{For_{>}}$, and ρ:

$$(G, F) \in \mathcal{D} \longrightarrow (\forall i \in W^G)\big((G, \rho), i \nVdash A > B \longrightarrow (F, \rho), i \nVdash A > B\big)$$

Proof : We have the following from *Proposition 4.5*, for all $F, G \in \mathbf{CS}$, $A, B \in \overline{For_{>}}$, and ρ:

1. $(F, G) \in \mathcal{R} \longrightarrow (\forall i \in W^F)\big((F, \rho), i \Vdash A > B \longrightarrow (G, \rho), i \Vdash A > B\big)$

Contraposing the consequent yields:

2. $(F, G) \in \mathcal{R} \longrightarrow (\forall i \in W^F)\big((G, \rho), i \nVdash A > B \longrightarrow (F, \rho), i \nVdash A > B\big)$

Finally, we obtain 3 by substituting an equivalent term in the antecedent of 2, by *Lemma 4.3*,

3. $(G, F) \in \mathcal{D} \longrightarrow (\forall i \in W^G)\big((G, \rho), i \nVdash A > B \longrightarrow (F, \rho), i \nVdash A > B\big)$

and note that whenever the antecedents of 2 and 3 are true, then $W^F = W^G$ is true, and the consequents of 2 and 3 are identical. If the antecedents of 2 and 3 are false, then both 2 and 3 are vacuously true, so the quantifier change is justified. □

4.4 Contextualized counterfactuals

4.4.1 Context representation

In order to account for contextualized counterfactuals, the formal language will be modified to include a family of connectives indexed by contexts. For that purpose, we need to posit the existence of an appropriate context (index) set. To sketch the background of the motivation for this approach, consider Lewis' (1981, p.218, my emphasis) view on the role of ordering frames:

> The ordering that gives the factual background depends on the facts about the world, known or unknown; how it depends on them is determined – or underdetermined – by our linguistic practice and by context. We may separate the contribution of practice and context from the contribution of the world, evaluating counterfactuals as true or false at a world, and according to *a 'frame' determined somehow by practice and context.*

In some sense what I am proposing is a little bit of a cheat, because prior to defining what contexts are I have already intended ordering frames to be the corresponding context-representations. That is, I have decided on a very precise character of representations of objects whose existence I merely stipulate. However, this correspondence is not only intuitive but also partly justified, since we have already shown that ordering frames can be

meaningfully interpreted as carriers of contextual information.[158] So, at least for our purpose of the minimal role of offering a means of distinguishing the truth of counterfactuals by explicit appeal to context in the object language – ordering frames serve as adequate model-theoretic candidates.

For an intuitive characterization of the proposed analysis and its capacity, let us proceed by examining some paradigmatic cases. Consider the following pairs of counterfactuals:

1. If Caesar had been in command, he would have used the atom bomb.
2. If Caesar had been in command, he would have used catapults.

3. If Everest was in New Zealand, Everest would be in the Southern Hemisphere.
4. If Everest was in New Zealand, New Zealand would be in the Northern Hemisphere.

Each counterfactual in the first pair can be true in some context, but rarely would we think of them both true in a single context (although that may be possible), and likewise not all contexts that make (1) true would make (2) true. That is, intuitively, each counterfactual can be true by virtue of distinct contextual assumptions. So, for instance (1) can be true in a context (call it context *a*) where Caesar's knowledge of weaponry is assumed to be that of a 20th century military general, and moreover that he would resort – the strategic and ruthless genius that he undoubtedly was – to the most effective means (available to him) of defeating the enemy. However (1) would hardly be true in a context (call it context *b*) where Caesar's knowledge of weaponry is restricted to that which he actually had in the 1st century BCE.

The current proposal meets this challenge and allows distinctions between the truth of counterfactuals with the same antecedent and consequent on any single occasion, by explicit appeal to contingent contextual considerations. For example, we could have (1) evaluated as true and false on the same occasion of utterance, since the object language (developed in the next section) allows for explicit reference to distinct contexts that influence the truth of (1), e.g. so (1.a) may be true, and (1.b) may be false, in particular if they're explicitly indexed by contexts *a* and *b*, characterized in the previous paragraph.

(1.a) In context *a* : If Caesar had been in command, he would have used the atom bomb.
(1.b) In context *b* : If Caesar had been in command, he would have used the atom bomb.

[158] Following Lewis (1973, §2.3; 1981, §2) in that regard.

By extension, the proposal accommodates truth differences and coincidences of counterfactuals with the same antecedent, but different consequents.[159] So, continuing our example, we could have (1) and (2) come apart in truth by appeal to a single context, i.e. have (1.b) remain false whilst having (2.b) correctly analysed as true.

(1.b) In context *b* : If Caesar had been in command, he would have used the atom bomb.
(2.b) In context *b* : If Caesar had been in command, he would have used catapults.

There may even be a single, peculiar context *c* where both (1) and (2) are true, however arguably there is no single context where both (3) and (4) are true, since once the enthymematic content is accounted for (all the relevant information is imported into the relevant antecedent worlds) we end up with inconsistency.[160] Now we turn to defining the contextualized language.

4.4.2 Modified languages

Each modified language is just like $\mathcal{L}$ given in *Definition 4.1.1* that generates For, but instead of the single connective $>$, each contains a family of indexed connectives.

Definition 4.4.1: Let $\mathcal{L}^{\mathcal{C}} := \{\sim, \Box, \Diamond, \wedge, \vee, \supset\} \cup \{>_c : c \in \mathcal{C}\}$, where $\mathcal{C}$ is a set, regarded as a set of contexts.

Here are some noteworthy special cases. Note that when $\mathcal{C} = \emptyset$, then $\mathcal{L}^{\mathcal{C}}$ is just the basic propositional modal language, and when $\mathcal{C}$ is a singleton, $\mathcal{L}^{\mathcal{C}}$ resembles $\mathcal{L}$ from *Definition 4.1.1* in the sense of being the basic propositional modal language expanded by a single additional dyadic modal connective.

Well-formed formulae will reflect the intended analysis, so context-indices will not vary across nested $>_c$-formulae. I propose that the context-index of the main conditional connective $>_c$ of a nested conditional, e.g. $A >_c (B >_c C)$ should settle the matter of what information is imported into counterfactual worlds when evaluating its subformulae. I do this

[159] Gabbay's (1972) account allows for this, but not for distinguishing in truth value counterfactuals with the same antecedent and consequent. Nute (1980, p.76) also makes this observation.
[160] This is because the consequents are not formulated in a manner as to suggest partial containment by either hemisphere, and the hemispheres are disjoint. I return to this example in §4.4.5 and give formal arguments demonstrating that we end up with inconsistency by granting the truth of both (3) and (4) in claims 4.5.2 and 4.5.3.

in *Definition 4.4.6* by stipulating that nested indexed-conditionals inherit the context-index of the outermost indexed conditional. This proposed approach goes *some way* of meeting the challenge posed by Priest's (2018, §3.1, f.14) question regarding what information from the world where the counterfactual is evaluated should be imported into counterfactual worlds, when evaluating nested conditionals (counterfactuals).

> An interesting question in this context is as follows. Consider a conditional with an embedded conditional, such as $A > (B > C)$. Is the information imported in evaluating the outer conditional the same as that imported in evaluating the inner conditional? […] Nothing said in this essay settles this matter.
> Priest (2018, §3.1, f.14)

The thought is that the information imported in evaluating the inner conditional is *contextually the same*, i.e. restricted by what information is imported in evaluating the outer conditional. The information is *not the same*, since the inner conditional need not have the same antecedent as the outer conditional, and its truth may not be evaluated at the same world as the outer conditional – both highly relevant factors that contribute to determining what information should be imported. The model theory given in §4.4.3 goes in detail how such contextual determination (of what information is imported) is established.

To define the set of well-formed formulae of interest, it will be easier to first define a larger set, and subsequently apply the required restrictions.

Definition 4.4.2: Let $for^{\mathcal{C}}$ be the smallest set closed under the following well-formed formula formation rules:

B: All propositional variables are wffs, i.e. $PV \subseteq for^{\mathcal{C}}$.

R1: If $A \in for^{\mathcal{C}}$ then $\{\sim A, \Box A, \Diamond A\} \subseteq for^{\mathcal{C}}$.

R2: If $\{A, B\} \subseteq for^{\mathcal{C}}$ then $\{A \wedge B, A \vee B, A \supset B\} \subseteq for^{\mathcal{C}}$.

R3: If $\{A, B\} \subseteq for^{\mathcal{C}}$ and $c \in \mathcal{C}$, then $A >_c B \in for^{\mathcal{C}}$

Definition 4.4.3: Let $\overline{For_{>}^{\mathcal{C}}}$ be the subset of $for^{\mathcal{C}}$, closed only under B, R1, and R2.
In other words, $\overline{For_{>}^{\mathcal{C}}}$ denotes the set of wffs of the basic modal language, i.e. it doesn't contain any instances of $>_c$ for any $c \in \mathcal{C}$.

Note that $\overline{For_{>}^{\mathcal{C}}} = \overline{For_{>}^{\mathcal{C}'}}$ for any two context sets $\mathcal{C}$ and $\mathcal{C}'$.

Definition 4.4.4: The *set of subformulae of* $A \in for^{\mathcal{C}}$ is the smallest set $Sub(A)$ satisfying the following conditions:

1. $A \in Sub(A)$
2. For each $* \in \{\sim, \Box, \Diamond\}$ if $* B \in Sub(A)$, then $B \in Sub(A)$.
3. For each $\circ \in \{\wedge, \vee, \supset\}$ if $B \circ C \in Sub(A)$, then $B \in Sub(A)$ and $C \in Sub(A)$.
4. For each $\circ \in \{>_c : c \in \mathcal{C}\}$ if $B \circ C \in Sub(A)$, then $B \in Sub(A)$ and $C \in Sub(A)$.

Definition 4.4.5: Let $Ind(\Sigma)$ be the set of context-indices appearing in $\Sigma \subseteq for^{\mathcal{C}}$.

$$Ind(\Sigma) = \{c \in \mathcal{C}: \exists A \in \Sigma \wedge [\exists B, C \in for^{\mathcal{C}}((B >_c C) \in Sub(A))]\}$$

Example: $Ind(\{p >_a (q >_d r), (p >_b r) \vee (p >_c q)\}) = \{a, b, c, d\}$.

Indexed conditionals embedded (nested) within other indexed conditionals inherit the indices of the outermost indexed conditional. It just doesn't make sense in this picture to speak of embedded conditionals whose indices vary. Below is the restriction on $for^{\mathcal{C}}$ that reflects this.

Definition 4.4.6: Let $For^{\mathcal{C}} := \{A \in for^{\mathcal{C}}: \forall c \in \mathcal{C}\left((C >_c D \in Sub(A)) \rightarrow Ind(\{C, D\}) \subseteq \{c\}\right)\}$.

Example: Formulae such as $p >_a (q >_b r)$ or $(q >_b r) >_a p$, where $a \neq b$, are not elements of $For^{\mathcal{C}}$. However, the following are: $p >_a (q >_a r)$, $(q >_b r) >_b p$, $(p >_a q) \vee (r >_b s)$.

Definition 4.4.7: Let $For^{\mathcal{C}}_{>0} := \{A \in For^{\mathcal{C}}: B >_c C \in Sub(A) \rightarrow B, C \in \overline{For^{\mathcal{C}}_{>}}\}$.

Example: $\sim(p >_a (q \supset r)) \wedge \left(((p \wedge \sim q) >_b r) \vee (q >_c r)\right) \in For^{\mathcal{C}}_{>0}$ but $p >_c (p >_c p) \notin For^{\mathcal{C}}_{>0}$ for any $a, b, c \in \mathcal{C}$.

Definition 4.4.8: Define $For^{\mathcal{C}}(>) := \{A >_c B: A, B \in For^{\mathcal{C}}, c \in \mathcal{C}\}$. That is, $For^{\mathcal{C}}(>)$ is just the set of $For^{\mathcal{C}}$ formulae whose main connective is $>_c$, for some $c \in \mathcal{C}$.

The following definition will play a key role in the definition of truth conditions for indexed counterfactuals, i.e. for truth conditions formulae like $A >_c B$.

Definition 4.4.9: Let $\_: For^{\mathcal{C}} \longrightarrow For$ be the function that transforms all formulae with indexed connectives $>_c$ for any $c \in \mathcal{C}$ into unindexed ones $>$, in all subformulae of a formula. That is, it "strips" any $For^{\mathcal{C}}$ formula of its indices leaving its index-less For counterpart.

B: $\underline{p} = p$ for all $c \in PV$.

R1: $\underline{* A} = * \underline{A}$ for each $* \in \{\sim, \Box, \Diamond\}$ and $A \in For^{\mathcal{C}}$.

R2: $\underline{A \circ B} = \underline{A} \circ \underline{B}$ for each $\circ \in \{\wedge, \vee, \supset\}$ and $A, B \in For^{\mathcal{C}}$.

R3: $\underline{A >_c B} = \underline{A} > \underline{B}$ for each $c \in \mathcal{C}$ and $A, B \in For^{\mathcal{C}}$.

Example: $\underline{\sim p >_c (q \vee r)} = \sim p > (q \vee r)$.

It will be useful to extend the above definition to *sets* of formulae. No ambiguity should arise whether the argument is a formula or a set of formulae.

Definition 4.4.10: For any $\Sigma \subseteq For^{\mathcal{C}}$, let $\underline{\Sigma} := \{\underline{A} \in For : A \in \Sigma\}$.

4.4.3 Modified model theory

The formula $A >_c B$ is intended to be read as explicitly contextualized version of $A > B$. That is, the model theory in this section provides an analysis of $A >_c B$, which is to be read as:

> 'In context *c*: If it were the case that A, then it would be the case that B.'

For the purposes of the modified model theory, we will need sets containing **CS** frames with some particular domain W (our models make use of **CS** models with the same domain).

Definition 4.4.11 : Let $\mathcal{F}_W := \{(U, \lesssim) \in \mathbf{CS} : U = W\}$.

Definition 4.4.12: A **CS**+ frame of the modified language is the triple:

$$(W, \mathcal{C}, r)$$

where $W \neq \emptyset$ and $\mathcal{C}$ are sets, $r : \mathcal{C} \longrightarrow \mathcal{F}_W$ is a *function*, and $\mathcal{F}_W$ is as defined in *4.4.11*. Informally, $\mathcal{C}$ is regarded as a set of contexts and $r_c \in r[\mathcal{C}] \subseteq \mathcal{F}_W$ is regarded as representing context *c*. Reflecting the earlier quote from Lewis r_c is the "ordering frame determined somehow by practice and context *c"*.

Additional restrictions, such as surjectivity and/or injectivity may be placed on *r*, to suit the preferred intended properties that a context representation function should have.

Definition 4.4.13.1: A **CS**+ model of the modified language is the quadruple:

$$(W, \mathcal{C}, r, \rho)$$

Where $(W, \mathcal{C}, r)$ is a **CS**+ frame and ρ is as in *Definition 4.2.5.*

Definition 4.4.13.2: Truth in **CS**+ models is defined in terms the satisfiability relation $\Vdash^{\mathcal{C}} \subseteq W \times For^{\mathcal{C}}$. We read $i \Vdash^{\mathcal{C}} A$ as 'A is true at i'.
When we want to explicitly refer to truth at a world in a particular **CS**+ model $\mathfrak{A}$, we shall employ the following notation: $\mathfrak{A}, i \Vdash^{\mathcal{C}} A$ and $\mathfrak{A}, i \Vdash^{\mathcal{C}} \Sigma$, as we have done for $\Vdash$.

Given a **CS**+ model $(W, \mathcal{C}, r, \rho)$ any $i \in W$, and any $c \in \mathcal{C}$ define $\Vdash^{\mathcal{C}}$ as follows:

(1) $i \Vdash^{\mathcal{C}} p$ iff $p\rho_i 1$
(2) $i \Vdash^{\mathcal{C}} \sim A$ iff not $i \Vdash^{\mathcal{C}} A$
(3) $i \Vdash^{\mathcal{C}} A \wedge B$ iff $i \Vdash^{\mathcal{C}} A$ and $i \Vdash^{\mathcal{C}} B$
(4) $i \Vdash^{\mathcal{C}} A \vee B$ iff $i \Vdash^{\mathcal{C}} A$ or $i \Vdash^{\mathcal{C}} B$
(5) $i \Vdash^{\mathcal{C}} A \supset B$ iff $i \Vdash^{\mathcal{C}} \sim A$ or $i \Vdash^{\mathcal{C}} B$
(6) $i \Vdash^{\mathcal{C}} \Box A$ iff $\forall j \in W: j \Vdash^{\mathcal{C}} A$.
(7) $i \Vdash^{\mathcal{C}} \Diamond A$ iff $\exists j \in W: j \Vdash^{\mathcal{C}} A$.
(8) $i \Vdash^{\mathcal{C}} A >_c B$ iff $(r_c, \rho), i \Vdash \underline{A >_c B}$

What's going on in (8)? The truth conditions for a formula $(A >_c B)$, i.e. with an indexed connective as the main connective, in a **CS**+ model are defined in terms of truth conditions for the corresponding non-indexed formula $(A > B)$ in a **CS** model – a model based on the ordering frame that is the image of c under r i.e. the ordering frame that is said to represent the context corresponding to the index of the indexed formula. This is how we formally capture the idea of indexed formulae being evaluated in contexts (represented by ordering frames) corresponding to the context index. Note that reference to a **CS** model is not required in any of the other clauses (1)-(8), since there are no formulae with an indexed connective as the main connective contained in the definienda of clauses (1)-(7).[161]

[161] An alternative, and semantically equivalent formulation of the truth conditions for the contextualized language would be to have r assigning comparative similarity assignments to worlds directly, relative to some context, i.e. to have $r: W \times \mathcal{C} \longrightarrow \wp(W) \times \wp(W \times W)$ be the function such that $r(w, c)$ is a comparative similarity assignment to world w, in context c. If we recall, this approach is closely aligned with Nolan's suggestion, highlighted at the end of chapter 1. But I chose not to go this way, since we can accommodate the contextual variability in **CS**+ models by recycling the formalism *already present* in **CS** ordering frames. Recall from the definition of **CS** ordering frames, that we already have defined a function $\lesssim: W \longrightarrow \wp(W) \times \wp(W \times W)$, which uniquely characterizes each ordering frame and whose image, for each world, consists of comparative similarity assignments being defined on $\wp(W) \times \wp(W \times W)$.

- As in the case of **CS** models, let's introduce the following notation for convenience:

$$i \Vdash^{\mathcal{C}} \Sigma \quad \text{iff} \quad i \Vdash^{\mathcal{C}} A \text{ for all } A \in \Sigma$$

- Also denote with $\mathfrak{A} \Vdash^{\mathcal{C}} A$ when $\mathfrak{A}, i \Vdash^{\mathcal{C}} A$ for all $i \in W^{\mathfrak{A}}$.

Note that it follows from the above definition that formulae whose index set ranges over more than one index may be evaluated on more than one **CS** model, e.g.

$(W, \mathcal{C}, r, \rho), i \Vdash^{\mathcal{C}} (A >_a B) \vee (C >_b D)$

iff $(W, \mathcal{C}, r, \rho), i \Vdash^{\mathcal{C}} (A >_a B)$ or $(W, \mathcal{C}, r, \rho), i \Vdash^{\mathcal{C}} (B >_b C)$

iff $(r_a, \rho), i \Vdash \underline{A >_a B}$ or $(r_b, \rho), i \Vdash \underline{C >_b D}$

That is, (r_a, ρ) and (r_a, ρ) are **CS** models, by definition, and they need not be the same.

Just as we have relativized formula validity to a model $\mathfrak{A} \Vdash^{\mathcal{C}} A$ it will be of use to define valid inference relativized to a model.

Definition 4.4.13.3: Let $\vDash^{\mathcal{C}}_{\mathfrak{A}} \subseteq \wp(For^{\mathcal{C}}) \times For^{\mathcal{C}}$, and given a **CS**+ model $\mathfrak{A} = (W, \mathcal{C}, r, \rho)$ write

- $\vDash^{\mathcal{C}}_{\mathfrak{A}} A$ iff $\mathfrak{A} \Vdash^{\mathcal{C}} A$
- $\Sigma \vDash^{\mathcal{C}}_{\mathfrak{A}} A$ iff for all $i \in W$: if $\mathfrak{A}, i \Vdash^{\mathcal{C}} \Sigma$, then $\mathfrak{A}, i \Vdash^{\mathcal{C}} A$.

Since each context set $\mathcal{C}$ gives rise to a distinct language $\mathcal{L}^{\mathcal{C}}$, and consequently a distinct set of wffs $For^{\mathcal{C}}$, we need a semantic consequence relation for each language. The definition below is of semantic consequence for each $\mathcal{L}^{\mathcal{C}}$. In most cases however, I'll omit the superscript $\mathcal{C}$ unless the discussion will hinge on some specific property of the context set.

Definition 4.4.14: Given a set $\mathcal{C}$ let $\vDash^{\mathcal{C}}_{\mathbf{CS+}} \subseteq \wp(For^{\mathcal{C}}) \times For^{\mathcal{C}}$, and define:

$\Sigma \vDash^{\mathcal{C}}_{\mathbf{CS+}} A$ iff for all **CS**+ models $\mathfrak{A}$ and $i \in W$: if $\mathfrak{A}, i \Vdash^{\mathcal{C}} B$ for all $B \in \Sigma$, then $\mathfrak{A}, i \Vdash^{\mathcal{C}} A$.

We say an inference from Σ to A is **CS**+ valid iff $\Sigma \vDash^{\mathcal{C}}_{\mathbf{CS+}} A$. That is, valid inference is defined as truth preservation at all worlds in all **CS**+ models. A formula $A \in For^{\mathcal{C}}$ is said to be **CS**+ valid iff $\emptyset \vDash^{\mathcal{C}}_{\mathbf{CS+}} A$. Call this logic (schema) **CS**+.

I use the term 'logic *schema*' since if $\mathcal{C} \neq \mathcal{C}'$, in particular if $|\mathcal{C}| \neq |\mathcal{C}'|$, then $\vDash^{\mathcal{C}}_{\mathbf{CS+}} \neq \vDash^{\mathcal{C}'}_{\mathbf{CS+}}$ by definition. For example, see *Corollaries 4.9.1* and *4.9.2*.

Note that it is immediate from the above definitions that $\vDash^{\mathcal{C}}_{\mathbf{CS+}} \subseteq \vDash^{\mathcal{C}}_{\mathfrak{A}}$, for any **CS**+ model $\mathfrak{A}$.

With the aid of the notation from *Definition 4.4.13.3* we can express **CS**+ semantic consequence definition more succinctly: $\Sigma \vDash^{\mathcal{C}}_{\mathbf{CS+}} A$ iff for all **CS+** models $\mathfrak{A}$: $\Sigma \vDash^{\mathcal{C}}_{\mathfrak{A}} A$.

Note that since the truth conditions for □ and ◊ formulae are defined in terms of unrestricted quantification over possible worlds, i.e. only $>_c$-formulae truth conditions depend on $\mathcal{C}$ and r, the above validity conditions give the modal logic **S5** for the basic modal language. This allows us to formulate a more precise statement about a special case, when $\mathcal{C}$ is empty.

Corollary 4.9.1: If $\mathcal{C} = \emptyset$, then $\Sigma \vDash^{\mathcal{C}}_{\mathbf{CS+}} A$ iff $\Sigma \vDash_{\mathbf{S5}} A$.
Proof : This follows immediately from the fact that if $\mathcal{C} = \emptyset$, then by *Definition 4.4.1* $\mathcal{L}^{\mathcal{C}}$ becomes $\{\sim, \Box, \Diamond, \wedge, \vee, \supset\} \cup \{>_c : c \in \emptyset\} = \{\sim, \Box, \Diamond, \wedge, \vee, \supset\}$, i.e. the basic modal language. □

There is another special case with interesting properties, when $\mathcal{C}$ is a singleton, which is expressed in *Corollary 4.9.2* at the beginning of the next section, shortly after *Theorem 4.9*.

The part of the basic modal language is indistinguishable between the two classes of models in the following sense.

Lemma 4.6: For any **CS**+ model $\mathfrak{M} = (W, \mathcal{C}, r, \rho)$ and any $A \in \overline{For^{\mathcal{C}}_{>}}$, $F \in \mathcal{F}_W$, $i \in W$:

$$\mathfrak{M}, i \Vdash^{\mathcal{C}} A \quad \text{iff} \quad (F, \rho), i \Vdash A$$

Proof : It suffices to note that elements of $\overline{For_{>}}$ depend only on W and ρ, which are the same for $\mathfrak{M}$ and (F, ρ), by definition. □

Theorem 4.7: If $\Sigma \cup \{A\} \subseteq \overline{For^{\mathcal{C}}_{>}}$: then $\Sigma \vDash_{\mathbf{CS}} A$ iff $\Sigma \vDash_{\mathbf{CS+}} A$.
Proof : Immediate from *Lemma 4.6*. □

Definition 4.4.15: Call frame $H \in \mathbf{CS}$ a *mutual refinement* of frames F and G iff $(F, H) \in \mathcal{R}$ and $(G, H) \in \mathcal{R}$. Note that H is a mutual refinement of F and G iff $H \in \mathcal{R}[F] \cap \mathcal{R}[G]$.[162]

It will be worthwhile (useful later) emphasizing a relatively obvious, yet important fact.

Lemma 4.8: If $(W, \lesssim) = F \in \mathbf{CS}$, then $\mathcal{R}[F] \subseteq \mathcal{F}_W$.
Proof : Immediate from definition of $\mathcal{F}_W$ and the fact that refinements preserve domains. □

[162] For a reminder of the meaning of $\mathcal{R}$ and $\mathcal{R}[F]$, see *definitions 4.3.1* and *4.3.1.2*, respectively.

4.4.4 Results

Much of Lewis' analysis is preserved on this account. This occurs when the premises and conclusion of an inference are confined to a single context. This makes sense intuitively, and the semantics manages to align with our intuition in this regard.

Theorem 4.9: For all $\Sigma \cup \{A\} \subseteq For^{\mathcal{C}}$:

If (1) $\underline{\Sigma} \vDash_{\mathbf{CS}} \underline{A}$ and

(2) $|Ind(\Sigma \cup \{A\})| \leq 1$,

then $\Sigma \vDash_{\mathbf{CS+}} A$.

In other words, (1) if the unindexed inference is **CS** valid, and (2) if the premises and conclusion range over at most one context-index, then the inference is **CS**+ valid.
Before we proceed, note that if $Ind(\Sigma \cup \{A\}) = \emptyset$, then the result follows from *Theorem 4.7*.

Proof : Let $\mathfrak{A} = (W, \mathcal{C}, r, \rho) \in$ **CS**+, such that $\mathfrak{A}, i \Vdash^{\mathcal{C}} B$ for all $B \in \Sigma$. We need to show that $\mathfrak{A}, i \Vdash^{\mathcal{C}} A$. Now, there is a $c \in \mathcal{C}$ such that for each $B \in \Sigma$, either $Ind(\{B\}) = \{c\}$, or $Ind(\{B\}) = \emptyset$, from (ii). If $Ind(\{B\}) = \{c\}$, then $(r_c, \rho), i \Vdash \underline{B}$, by *Definition 4.4.13.2*. Otherwise, if $Ind(\{B\}) = \emptyset$, then $(F, \rho), i \Vdash \underline{B}$, for all $F \in \mathcal{F}_W$, by *Lemma 4.6*, and we note that $r_c \in \mathcal{F}_W$. Hence, $(r_c, \rho), i \Vdash \underline{B}$ for all $B \in \Sigma$. Hence $(r_c, \rho), i \Vdash \underline{A}$, by (i). Hence, $\mathfrak{A}, i \Vdash^{\mathcal{C}} A$, by *Definition 4.4.13.2*, as required. □

Recalling how $For^{\mathcal{C}}$ has been defined, i.e. that all nested counterfactuals inherit the index of the outermost counterfactual, *Theorem 4.9* sanctions a number of important inference patterns.

Example: For all $\mathcal{C} \neq \emptyset$, $A, B, C \in For^{\mathcal{C}}$, and all $c \in \mathcal{C}$:

$\vDash_{\mathbf{CS+}} A >_c A$

$A, A >_c B \vDash_{\mathbf{CS+}} B$

$\sim B, A >_c B, \vDash_{\mathbf{CS+}} \sim A$

$A, B \nvDash_{\mathbf{CS+}} A >_c B$

$\vDash_{\mathbf{CS+}} (A \wedge \sim A) >_c B$

$\Box(A \supset B) \vDash_{\mathbf{CS+}} A >_c B$

The results follow directly from *Theorem 4.9* and the definition of $For^{\mathcal{C}}$. Clearly, **CS**+ inherits the vacuous treatment of counterpossibles from **CS** (as will all the systems based on **CS** models).

We have looked earlier at the special case of **CS**+, when $\mathcal{C} = \emptyset$, and shown in *Corollary 4.9.1* that **CS**+ reduces to **S5**, i.e. $\vDash^{\emptyset}_{\mathbf{CS+}} = \vDash_{\mathbf{S5}}$. There is another important special case, when $\mathcal{C}$ is a singleton, and a corresponding result, of **CS**+ reducing to **CS**, i.e. essentially $\vDash^{\{\emptyset\}}_{\mathbf{CS+}} = \vDash_{\mathbf{CS}}$.

Corollary 4.9.2: If $|\mathcal{C}| = 1$, then: $\underline{\Sigma} \vDash_{\mathbf{CS}} \underline{A}$ iff $\Sigma \vDash^{\mathcal{C}}_{\mathbf{CS+}} A$

Proof : (⟶) Is just *Theorem 4.9* because if $\mathcal{C}$ is a singleton then condition (2) is always satisfied. (⟵) It suffices to observe that since *r* is a function, the image of $\mathcal{C}$ under *r* for each **CS**+ frame is also a singleton, i.e. $r[\mathcal{C}] = \{F\}$ for each **CS**+ frame $(W, \mathcal{C}, r)$, where F is a **CS** frame by definition of *r*. Suppose for contradiction that $\Sigma \vDash^{\mathcal{C}}_{\mathbf{CS+}} A$ and $\underline{\Sigma} \nvDash_{\mathbf{CS}} \underline{A}$. From $\underline{\Sigma} \nvDash_{\mathbf{CS}} \underline{A}$ we infer that there's a **CS** model $\mathfrak{A} = (F, \rho^{\mathfrak{A}})$ and $i \in W^{\mathfrak{A}}$ such that $(F, \rho^{\mathfrak{A}}), i \Vdash \underline{\Sigma}$ and $(F, \rho^{\mathfrak{A}}), i \nVdash \underline{A}$. But consider the **CS**+ model $\mathfrak{B} = (\mathfrak{F}, \rho^{\mathfrak{B}})$ such that $W^{\mathfrak{A}} = W^{\mathfrak{B}}$, $\rho^{\mathfrak{A}} = \rho^{\mathfrak{B}}$, and $r^{\mathfrak{B}}[\mathcal{C}] = \{F\}$, i.e. in $\mathfrak{B}$ all indexed formulae are evaluated on $(F, \rho^{\mathfrak{A}}) = \mathfrak{A}$. But $(F, \rho^{\mathfrak{A}}), i \Vdash \underline{\Sigma}$ implies $(\mathfrak{F}, \rho^{\mathfrak{B}}) \Vdash^{\mathcal{C}} \Sigma$, by definition of $\Vdash^{\mathcal{C}}$. But then $(\mathfrak{F}, \rho) \Vdash^{\mathcal{C}} A$ by hypothesis $\Sigma \vDash^{\mathcal{C}}_{\mathbf{CS+}} A$, which implies $(F, \rho^{\mathfrak{A}}), i \Vdash \underline{A}$ by definition of $\Vdash^{\mathcal{C}}$, which contradicts $(F, \rho^{\mathfrak{A}}), i \nVdash \underline{A}$. □

Naturally, the advantages of **CS**+ appear when $|\mathcal{C}| > 1$.

The main application of our key result, about frame refinements, i.e. *Proposition 4.5*, is in condition (2) of the following theorem. The theorem is the second major step in developing a notion of contextualized inference in the form of systems **CS1**+ and **CS2**+, defined in the next section. In particular it establishes an important relationship between the contextual information carried by the premises and the conclusion.

Note that the restriction of $For_{>}$ to $For^{\mathcal{C}}_{>0} \cap For^{\mathcal{C}}(>)$ stems from the fact that ordering frame refinements are only truth preserving, and that's the part of $For_{>}$ to which *Proposition 4.5* applies. Just to be clear, if $For^{\mathcal{C}}_{>0} \cap For^{\mathcal{C}}(>) \neq \emptyset$, it contains only formulae $A >_c B$ such that $A, B \in \overline{For^{\mathcal{C}}_{>}}$. That is $A \in For^{\mathcal{C}}_{>0} \cap For^{\mathcal{C}}(>)$ iff $|Ind(\{A\})| = 1$. In other words, this result apples to a language restricted to the basic propositional modal language with indexed conditionals appearing only as the main connectives to formulae (i.e. that are not a proper subformula of any formula) that don't contain any other indexed conditionals as proper subformulae.

Theorem 4.10: For all $\Sigma \cup \{A\} \subseteq \left(For^{\mathcal{C}}_{>0} \cap For^{\mathcal{C}}(>)\right) \cup \overline{For^{\mathcal{C}}_{>}}$:

If (1) $\underline{\Sigma} \vDash_{\mathbf{CS}} \underline{A}$ and

(2) for each **CS**+ model:

(i) if $Ind(\{A\}) = \emptyset$, then $\bigcap\{\mathcal{R}[r_b] : b \in Ind(\Sigma)\} \neq \emptyset$, and

(ii) if $|Ind(\{A\})| = 1$, then $r_a \in \bigcap\{\mathcal{R}[r_b]\colon b \in Ind(\Sigma)\}$ for $\{a\} = Ind(\{A\})$,

then $\Sigma \vDash_{\mathbf{CS+}} A$.

In other words, (1) if the unindexed inference is **CS** valid, and (2) if the frame representation of the conclusion context-index is a mutual refinement of frame representations of context indices over which the premises range, then the inference is **CS**+ valid. We interpret condition (2) as saying that the context on which the conclusion is evaluated is not independent of the contexts on which the premises are evaluated, i.e. the conclusion is evaluated on an ordering frame that *preserves the contextual information* carried by ordering frames on which the premises are evaluated. Before we proceed with the proof, note that if $Ind(\Sigma \cup \{A\}) = \emptyset$, then the result follows from *Theorem 4.7*.

Proof : Let $\mathfrak{A} = (W, \mathcal{C}, r, \rho) \in$ **CS**+, such that $\mathfrak{A}, i \Vdash^{\mathcal{C}} B$ for all $B \in \Sigma$. We need to show that $\mathfrak{A}, i \Vdash^{\mathcal{C}} A$. Now, for each $B \in \Sigma$, and any $b \in \mathcal{C}$, if $Ind(\{B\}) = \{b\}$, then $(r_b, \rho), i \Vdash \underline{B}$, by *Definition 4.4.13.2*. Else if $Ind(\{B\}) = \emptyset$, then $(F, \rho), i \Vdash \underline{B}$ for all $F \in \mathcal{F}_W$, by *Lemma 4.6*, and we note that $\cup\{\mathcal{R}[r_b]\colon b \in Ind(\Sigma)\} \subseteq \mathcal{F}_W$. Suppose $|Ind(\{A\})| \leq 1$. First, suppose $|Ind(\{A\})| = 0$, and infer from (2.i) that there is a **CS** frame $G \in \mathcal{F}_W$ such that $G \in \bigcap\{\mathcal{R}[r_b]\colon b \in Ind(\Sigma)\}$. Next, we have $(r_b, \rho), i \Vdash \underline{B} \implies (G, \rho), i \Vdash \underline{B}$ for each $B \in For^{\mathcal{C}}_{>0} \cap For^{\mathcal{C}}(>)$, $b \in Ind(\{B\})$, from *Proposition 4.5*. Hence, $(G, \rho), i \Vdash \underline{B}$ for all $B \in \Sigma$ such that $Ind(\{B\}) \neq \emptyset$. From *Lemma 4.6* and $G \in \mathcal{F}_W$ we infer $(G, \rho), i \Vdash \underline{B}$ for each $B \in \Sigma$ such that $Ind(\{B\}) = \emptyset$. Hence, $(G, \rho), i \Vdash \underline{B}$ for all $B \in \Sigma$. Hence $(G, \rho), i \Vdash \underline{A}$, by (1). Hence $\mathfrak{A}, i \Vdash^{\mathcal{C}} A$, by *Lemma 4.6*. Next, suppose $|Ind(\{A\})| = 1$. From (2.ii) we have $r_a \in \bigcap\{\mathcal{R}[r_b]\colon b \in Ind(\Sigma)\}$ for any $a \in \mathcal{C}$ such that $Ind(\{A\}) = \{a\}$. By letting $r_a = G$, the reminder of the proof continues by the same reasoning as in the previous case, to the point where we conclude that $(r_a, \rho), i \Vdash \underline{A}$. From there, we conclude $\mathfrak{A}, i \Vdash^{\mathcal{C}} A$, by *Definition 4.4.13.2*, as required. □

In particular, the **CS**-validity of *Adjunction of Consequents* is preserved. This inference form will serve as a guiding example in the next section, motivating the reformulation of the current definition of **CS**+ valid inference.

Corollary 4.10.1: For all $A, B, C \in \overline{For^{\mathcal{C}}_{>}}$, and all $a, b, c \in \mathcal{C}$:
If $r_c \in \mathcal{R}[r_a] \cap \mathcal{R}[r_b]$ for all **CS+** models, then $A >_a B, A >_b C \vDash_{\mathbf{CS+}} A >_c (B \wedge C)$.

Proof : The result follows directly from *Theorem 4.10*. □

Corollary 4.10.2: For all $A, B, C \in \overline{For_{>}^{\mathcal{C}}}$, and all $a, b, c, d \in \mathcal{C}$:
If $r_d \in \mathcal{R}[r_a] \cap \mathcal{R}[r_b] \cap \mathcal{R}[r_c]$ for all **CS+** models, then $A >_a B, B >_b A, A >_c C \vDash_{\mathbf{CS+}} B >_d C$.

Proof : The result follows directly from *Theorem 4.10*. □

4.4.5 Contextualized validity: discussion

CS+ is very weak since on the current definition 4.4.14 of **CS**+ valid inference there are no additional conditions placed on the relationship between context-indices appearing in the premises and the conclusion. But this is inadequate if we wish to fashion a logic that is sensitive to explicit contextual content. That is, we have developed an analysis of the contextualized language but have only included *truth preserving* conditions for validity in that definition – naturally, we also want a notion of *contextual information preserving* conditions on the new, contextualized notion of valid inference.

That is, currently, by *Definition 4.4.14* we have the following condition for **CS**+ valid inference:

$$\Sigma \vDash_{\mathbf{CS+}}^{\mathcal{C}} A \quad \text{iff} \quad \Sigma \vDash_{\mathfrak{A}}^{\mathcal{C}} A \text{ for all } \mathbf{CS}\text{+ models } \mathfrak{A}.$$

Where $\Sigma \vDash_{\mathfrak{A}}^{\mathcal{C}} A$ is: $\Sigma \vDash_{\mathfrak{A}}^{\mathcal{C}} A$ iff for all $i \in W$: if $\mathfrak{A}, i \Vdash^{\mathcal{C}} \Sigma$, then $\mathfrak{A}, i \Vdash^{\mathcal{C}} A$, as in *Def. 4.4.13.3*.

Clearly, these validity conditions are no different from those for **CS**. Such conditions make **CS**+ much weaker than **CS**, because for every **CS** valid inference there will be a counterexample by choice of indices for the premises and conclusion such that the premises are true, and the conclusion is false.

4.4.5.1 Contextualized validity: system CS1+

Theorem 4.10 captures some of the contextual information preserving features that hint at how contextual constraints could be fashioned. The theorem tells us that if we restrict the language in a way that *Proposition 4.5* can be implemented, then **CS** validity and valid inference is preserved if additional conditions on the relationship between the premises index set and conclusion index are satisfied, i.e. conditions that correspond to what we mean by contextual information preservation. This opens a possibility of defining a notion of valid inference that those conditions underlie. That is, as our initial attempt, we could fashion a notion of contextualized inference by adding condition (2) of *Theorem 4.10* to the current

definition **CS**+ validity and valid inference. The key definition that requires change is of $\Sigma \models^{\mathcal{C}}_{\mathfrak{A}} A$, since $\Sigma \models^{\mathcal{C}}_{\mathbf{CS+}} A$ is defined in terms of it.

Definitions 4.5.2 and 4.5.3 establish a proper logic of contextualized counterfactuals. That is, a logic where valid inference is not defined merely in terms of truth preservation but also in terms of contextual information preservation. Let us introduce some useful shorthand notation first.

Definition 4.5.1: Denote $For^{\mathcal{C}}_{>0} \cap For^{\mathcal{C}}(>)$ with $For^{\mathcal{C}}_{>0}(>)$.

Definition 4.5.2: Given a set $\mathcal{C}$ let $\models^{\mathcal{C}}_{\mathfrak{A}} \subseteq \wp\left(For^{\mathcal{C}}_{>0}(>) \cup \overline{For^{\mathcal{C}}_{>}}\right) \times \left(For^{\mathcal{C}}_{>0}(>) \cup \overline{For^{\mathcal{C}}_{>}}\right)$ and define for a **CS**+ model $\mathfrak{A}$, $\Sigma \models^{\mathcal{C}}_{\mathfrak{A}} A$ iff for all $\Sigma \cup \{A\} \subseteq For^{\mathcal{C}}_{>0}(>) \cup \overline{For^{\mathcal{C}}_{>}}$:

If for all $i \in W$: (i) $\mathfrak{A}, i \Vdash^{\mathcal{C}} \Sigma$, and

(ii) if $Ind(\{A\}) = \emptyset$, then $\bigcap\{\mathcal{R}[r_b] : b \in Ind(\Sigma)\} \neq \emptyset$, and
if $|Ind(\{A\})| = 1$, then $r_a \in \bigcap\{\mathcal{R}[r_b] : b \in Ind(\Sigma)\}$ for $\{a\} = Ind(\{A\})$,

then $\mathfrak{A}, i \Vdash^{\mathcal{C}} A$.

For formula validity, write $\models^{\mathcal{C}}_{\mathfrak{A}} A$ iff $\mathfrak{A}, i \Vdash^{\mathcal{C}} A$ for all $i \in W^{\mathfrak{A}}$, as given in *Definition 4.4.13.2*.

Now (model) validity is additionally conditioned on (ii) which intends to capture the idea that we evaluate the conclusion on a context that preserves contextual information of the contexts over which the premises range. Cases where (ii) is not satisfied will go through vacuously, thus disabling many counterexamples that would have been possible on *Definition 4.4.13.3* of $\models^{\mathcal{C}}_{\mathfrak{A}}$, which underlies **CS**+ validity. That is, it is no longer possible on the above definition to pick arbitrary indices for the premises and conclusion to generate counterexamples. Call the logic that satisfies this additional contextual information preservation constraint **CS1**+. It certainly is a step in the right direction, but one a little too far – the logic is too strong. As I'll shortly argue, it requires further finetuning, else it would give an incorrect analysis of a family of paradigmatic inference forms, i.e. by formally validating inference forms that are intuitively invalid.[163]

[163] By *paradigmatic* here I mean inference forms that can be said to emphasize the character of contextualized inference. That is, I have in mind the simplest inference forms whose validity turns on contextual considerations.

Definition 4.5.3: For all $\Sigma \cup \{A\} \subseteq For_{>0}^{\mathcal{C}}(>) \cup \overline{For_{>}^{\mathcal{C}}}$: write $\Sigma \models_{\mathbf{CS1+}}^{\mathcal{C}} A$ iff $\Sigma \models_{\mathfrak{A}}^{\mathcal{C}} A$ for all **CS**+ models $\mathfrak{A}$, where $\Sigma \models_{\mathfrak{A}}^{\mathcal{C}} A$ is as defined in 4.5.2. Also write $\models_{\mathbf{CS1+}}^{\mathcal{C}} A$ iff $\models_{\mathfrak{A}}^{\mathcal{C}} A$ for all **CS**+ models $\mathfrak{A}$. Call this logic (schema) **CS1**+.

4.4.5.2 Properties of CS1+

It should not be a surprise that **CS**+ validity and valid inference based on the above definition of $\Sigma \models_{\mathfrak{A}}^{\mathcal{C}} A$, in conjunction with *Theorem 4.10* yields the following.

Corollary 4.11: For all $\Sigma \cup \{A\} \subseteq For_{>0}^{\mathcal{C}}(>) \cup \overline{For_{>}^{\mathcal{C}}}$:

$$\text{If } \underline{\Sigma} \models_{\mathbf{CS}} \underline{A}, \text{ then } \Sigma \models_{\mathbf{CS1+}}^{\mathcal{C}} A.$$

Proof : The proof proceeds much like the proof of *Theorem 4.10*. We assume the antecedent (which is just condition (1) in *Theorem 4.10*), and for arbitrary **CS**+ model $\mathfrak{A}$ and $i \in W^{\mathfrak{A}}$ we assume (i) $\mathfrak{A}, i \Vdash^{\mathcal{C}} \Sigma$ and (ii) from *Definition 4.5.2* (where (ii) is just a special case of condition (2) in *Theorem 4.10*, relativized to $\mathfrak{A}$, and (i) is the starting hypothesis in the proof of *Theorem 4.10*) and then we proceed to show that $\mathfrak{A}, i \Vdash^{\mathcal{C}} A$, which is exactly what is to be shown in the proof of *Theorem 4.10*. □

4.4.5.3 CS1+ is too strong

As mentioned earlier **CS1**+ is too strong for the contextualized language, because it validates contextualized *Adjunction of Consequents*, despite some obvious counterexamples. In other words, by *Corollary 4.11*, we have $A >_a B, A >_b C \models_{\mathbf{CS1+}}^{\mathcal{C}} A >_c (B \wedge C)$ for all $A, B, C \in \overline{For_{>}^{\mathcal{C}}}$, and $a, b, c \in \mathcal{C}$, but as I'll argue we want it to fail. That is, we want to transform a family of instances of *Adjunction of Consequents* that go through vacuously on **CS1**+ (because of (ii) not being satisfied) into counterexamples. Since *Adjunction of Consequents* is **CS** valid, it is also **CS1**+ valid (by *Corollary 4.11*) but is seems clearly invalid for contextualized counterfactuals. I will now discuss instances of contextualized *Adjunction of Consequents* that I believe should be analysed as counterexamples but currently go through on **CS1**+ because of the absence of a mutual refinement of ordering frames corresponding to the premise context indices, and so *a fortiori* the ordering frame corresponding to the conclusion context index can't be such a mutual refinement. Consequently (ii) is rendered false and the questionable instance goes through vacuously. To outline the culminating point of the following discussion, let me just say at this point that the corresponding fix will be to

strengthen the validity conditions by adding a second condition that demands the existence of a mutual refinement of ordering frames corresponding to premise context indices, which itself corresponds to the conclusion context index.

Consider the following counterexample to *Adjunction of Consequents*, which is a limit case of the extent to which context can diverge, i.e. when the contextual information carried by ordering frames corresponding to the premise context-indices is incompatible.

Example 4.5.1: Counterexample to *Adjunction of Consequents*

(1) If Mt. Everest was in New Zealand, Everest would be in the Southern Hemisphere.

(2) If Mt. Everest was in New Zealand, New Zealand would be in the Northern Hemisphere.

(3) Therefore, if Mt. Everest was in New Zealand, then Mt. Everest would be in the Southern Hemisphere and New Zealand would be in the Northern Hemisphere.

The premises seem fine if taken separately – each in its own context (which the contextualized account allows), much like Quine's example with Caesar – but the conclusion is not only false, but absurd.[164] It should be regarded as a counterexample schema, since there are infinitely many examples like it, all of which speak against the validity of this inference form. Unfortunately, examples like 4.5.1 go through vacuously on of **CS1**+, because although **CS**+ models allow for the mutual truth of both premises, there is no mutual refinement of ordering frames corresponding to the context-indices of both premises (claim 4.5.3). There's no such refinement, since that would imply the existence of a **CS** model where both premises are true, which is impossible (claim 4.5.2). Both claims 4.5.2 and 4.5.3 are proven once the counterexample is sufficiently formalized, and the imported information highlighted.
Let us make the example formally precise and reveal all the imported information and relevant enthymemes. A translates to E in Z (**E**verest is in N**Z**), B translates to E in S (**E**verest is in the **S**outhern Hemisphere), and C translates to Z in N (N**Z** is in the **N**orthern Hemisphere. The enthymemes are: $N \cap S = \emptyset$ and 'none of E, N, S, Z is empty', i.e. the Northern and Southern hemispheres are disjoint, and all objects referred to explicitly have a nonzero spatial extension. Here we need to import in both cases, it seems, the information that $N \cap S = \emptyset$, i.e.

[164] Or as Priest (2017) would say 'both can be heard as true, but different information is imported in each case'.

it's part of the *ceteris paribus* clause (worlds where the hemispheres are disjoint are more similar to the actual world than worlds where they're not disjoint).

Counterexample with explicated details.

P.1	$(E \text{ in } Z) > (E \text{ in } S)$ [true]	(imported factual information: $Z \subseteq S$)
P.2	$(E \text{ in } Z) > (Z \text{ in } N)$ [true]	(imported factual information: $E \subseteq N$)
E.1	$N \cap S = \emptyset$ [true]	(relevant enthymeme)
E.2	None of E, N, S, Z is empty. [true]	(intended enthymeme)

———————————

∴	$E \text{ in } Z > (E \text{ in } S \wedge Z \text{ in } N)$ [false]

It should be noted that the notion of *information importation*, as described by Priest (2018,§2.1), and presented as *Definition 1.15* in chapter 1, is robust enough to be incorporated to our semantics – it is just the information that we import into the most similar antecedent worlds. It offers another way of talking about the worlds relevant to evaluating the corresponding material conditional when evaluating $A >_c B$.[165] The counterexample goes through, as mentioned earlier, because both premises can never be true, once we account for the relevant factual information imported into the antecedent worlds.[166]

4.4.5.4 *Adjunction of Consequents* – a comparative analysis

The differences in analyses are the following: On C, C+, and **CS** example 4.5.1 goes through vacuously, because both premises can't be true.[167] That is, such examples go through because once the relevant information is imported into the antecedent worlds the *combined truth of both counterfactual premises*, at some world *implies inconsistent situations* (at the relevant

[165] See *Definition 4.2.5*, and non-vacuous **CS** truth conditions for $A > B$, i.e. (8), (ii).
[166] It turns out that this example is a lot like Bennett's "East Gate, West Gate" formulation of an example given by Gibbard (1981). Priest (2017) gives an interesting analysis of this scenario. The scenario resembles example 4.5.1 in the sense that both premises are true in distinct contexts, but once the information required to make each premise true is jointly imported we get inconsistency, or implicit inconsistency.
[167] For definitions of C and C+ see §2.1.3 on *ceteris paribus* conditionals.

antecedent worlds). So, if the analysis is restricted to possible worlds, the premises can't be jointly true (since inconsistent situations can't be accommodated on possible worlds semantics), thus allowing the conclusion to follow vacuously. But it seems they can be jointly true, when taken in their appropriate (albeit distinct) contexts, which **CS**+ models permit. Note that if there is no **CS** model where both premises can be jointly true means that there is no **CS**+ model such that there would exist a mutual refinement of the ordering frames corresponding to premise context indices, since that would imply the existence of a **CS** model where the premises *can* be true, which is impossible (I'll return to this matter shortly).

NOTATION: for the purpose of the next few proofs, let us introduce some useful notation.

Definition 4.5.4: Let $(S, \lesssim)$ be a preordered set and $x \in S$ Define $\downarrow x_{(S,\lesssim)} := \{y \in S: y \lesssim x\}$. When there is no ambiguity regarding the preordered set in question, I'll omit the subscript.

The above definition lets us reformulate more succinctly the non-vacuous case of **CS** truth conditions for formulae expressing counterfactuals, i.e. we can rewrite (8).(ii) of *Definition 4.2.5*:

(8*) $i \Vdash A > B$ iff (i) $\sim\exists k \in S_i: k \Vdash A$, or
(ii) $\exists k \in W: k \Vdash A$ and $\downarrow k_i \cap [A] \subseteq [B]$.

Where $\downarrow k_i$ is just shorthand for $\downarrow k_{(W,\lesssim_i)}$. In contexts where the subscript is constant, I'll omit it altogether and just write $\downarrow k$ for brevity.

Presently I show that the inference $A > B, A > C \vDash A > (B \wedge C)$, viz. *Adjunction of Consequents* is valid on conditional logic C, and therefore on its extensions (notably C+), and **CS** and its extensions.

Proposition 4.12: $A > B, A > C \vDash_{\mathbf{CS}} A > (B \wedge C)$

Proof : Let $\mathfrak{A} = (W, \lesssim, \rho)$ be a **CS** model, and let $i \Vdash \{A > B, A > C\}$ for an arbitrary world $i \in W$. Then $\exists k \in W: k \Vdash A$ such that $\downarrow k \cap [A] \subseteq [B]$, and $\exists k' \in W: k' \Vdash A$ such that $\downarrow k' \cap [A] \subseteq [C]$. Now, either $k \lesssim_i k'$ or $k' \lesssim_i k$, by totality of $\lesssim_i$. If $k \lesssim_i k'$, then clearly $\downarrow k \subseteq \downarrow k'$, which in conjunction with the hypothesis implies that $\downarrow k \cap [A] \subseteq [C]$. Hence, we have both $\downarrow k \cap [A] \subseteq [B]$ and $\downarrow k \cap [A] \subseteq [C]$, which jointly imply $\downarrow k \cap [A] \subseteq [B] \cap [C]$. Hence $i \Vdash A > (B \wedge C)$, as required. A very similar argument holds for the case when $k' \lesssim_i k$. □

It's easy to show that *Adjunction of Consequents* is also valid on conditional logic C.[168]

Claim 4.5.2: Given *Example 4.5.1*, there's no **CS** model where both premises can be true. *Proof* : We have the key enthymeme $[A \text{ in } S] \cap [A \text{ in } N] = \emptyset$ for any extended area A on the surface of the Earth. Let's start with the first counterfactual. For $i \Vdash (E \text{ in } Z) > (E \text{ in } S)$ we require that $\exists k \in [E \text{ in } Z]$ and $\downarrow k \cap [E \text{ in } Z] \subseteq [E \text{ in } S]$, and accounting for imported information yields $\downarrow k \cap [E \text{ in } Z] \subseteq [E \text{ in } S] \cap [Z \text{ in } S]$, which implies (*) $\downarrow k \cap [E \text{ in } Z] \nsubseteq [Z \text{ in } N]$. Next, for the second premise $i \Vdash (E \text{ in } Z) > (Z \text{ in } N)$ we require $\exists k' \in [E \text{ in } Z]$ and $\downarrow k' \cap [E \text{ in } Z] \subseteq [Z \text{ in } N]$, which implies (**) $\downarrow k' \cap [E \text{ in } Z] \nsubseteq [Z \text{ in } S]$. Given totality, either $k \lesssim_i k'$ or $k' \lesssim_i k$. Hence, either $\downarrow k \subseteq \downarrow k'$ or $\downarrow k' \subseteq \downarrow k$. First suppose $\downarrow k \subseteq \downarrow k'$. Then, given that $\downarrow k' \cap [E \text{ in } Z] \subseteq [Z \text{ in } N]$ it follows that $\downarrow k \cap [E \text{ in } Z] \subseteq [Z \text{ in } N]$, which contradicts (*). Next, suppose $\downarrow k' \subseteq \downarrow k$. Then given that we have $\downarrow k \cap [E \text{ in } Z] \subseteq [Z \text{ in } S]$ it follows that $\downarrow k' \cap [E \text{ in } Z] \subseteq [Z \text{ in } S]$, which contradicts (**). So, if the first premise is true, the second premise can't be true. A very similar argument shows that if we assume the truth of the second premise, the first premise can't be true. □

It is easy to show that an analogous claim holds for logic C (and its extensions).[169]

A more intuitive way of seeing this, is to note that for $i \Vdash (E \text{ in } Z) > (E \text{ in } S)$ we require that all antecedent worlds where both Mt. Everest and NZ are in the Southern Hemisphere are more similar than worlds where both those objects are in the Northern Hemisphere, but for $i \Vdash (E \text{ in } Z) > (Z \text{ in } N)$ we require (the opposite) that all antecedent worlds where both Mt. Everest and NZ are in the Northern Hemisphere are more similar than worlds where both those objects are in the Southern Hemisphere. Both orderings are clearly incompatible. That is, for the first premise to be true there will be an antecedent world k_0 where both objects are in the Southern Hemisphere and all worlds j_{NH} where both objects are in the Northern Hemisphere satisfy (i): $\forall j(k_0 <_i j_{NH})$, whereas for the second premise to be true there will be

[168] $A > B, A > C \vDash_{\mathrm{C}} A > (B \wedge C)$. *Proof* : Let $\mathfrak{A} = (W, R, V)$ be a C model and let $i \Vdash \{A > B, A > C\}$ for arbitrary world $i \in W$. Then we have both $f_A(i) \subseteq [B]$ and $f_A(i) \subseteq [C]$, which implies that $f_A(i) \subseteq [B] \cap [C] = [B \wedge C]$. Therefore, $i \Vdash A > (B \wedge C)$, as required.

[169] Given *Example 4.5.1*, there is no C model where both premises can be true. *Proof* : First it needs to be granted that $f_{E \text{ in } Z}(@) \neq \emptyset$. Moreover, we have the key enthymematic fact $[A \text{ in } S] \cap [A \text{ in } N] = \emptyset$ for any area A on the surface of the Earth. Let's start with P.1. It's true at the actual world iff $f_{E \text{ in } Z}(@) \subseteq [E \text{ in } S]$, which implies that $f_{E \text{ in } Z}(@) \subseteq [E \text{ in } S] \cap [Z \text{ in } S]$, once we account for the imported information into $f_{E \text{ in } Z}(@)$, which implies $f_{E \text{ in } Z}(@) \subseteq [Z \text{ in } S]$, which implies $f_{E \text{ in } Z}(@) \nsubseteq [Z \text{ in } N]$. But $f_{E \text{ in } Z}(@) \subseteq [Z \text{ in } N]$ is necessary for the truth of P.2. So, if the first premise is true, the second premise can't be true. A very similar argument shows that if we assume the truth of the second premise, the first premise can't be true.

an antecedent world j_0 where both objects are in the Northern Hemisphere and all worlds k_{SH} where both objects are in the Southern Hemisphere satisfy (ii): $\forall k(j_0 <_i k_{SH})$. Now, $\forall j(k_0 <_i j_{NH})$ implies $k_0 <_i j_0$, and $\forall k(j_0 <_i k_{SH})$ implies $j_0 <_i k_0$, jointly yielding $k_0 <_i j_0$ and $j_0 <_i k_0$, which is impossible.

Claim 4.5.3: Given *Example 4.5.1*, it follows that for any **CS+** model where both premises are true there is no mutual refinement of the ordering frames corresponding to the premise context indices.

Informal proof : This is clear when we consider this as a corollary of *Claim 4.5.2*. That is, since $(E \text{ in } Z) > (E \text{ in } S)$ and $(E \text{ in } Z) > (Z \text{ in } N)$ can't be both true on any **CS** model, as we have established in *Claim 4.5.2*, then, in particular, for any **CS**+ model where the premises are both true, there is no mutual refinement of the ordering frames corresponding to the premise indices, since that would imply that both premises are true on some **CS** model (recall that refinements are truth preserving), which we have established in *Claim 4.5.2* as impossible. Below is a formal proof.

Proof : For any **CS**+ model $\mathfrak{A} = (W, \mathcal{C}, r, \rho)$, $i \in W$, and $a, b \in \mathcal{C}$, if $\mathfrak{A}, i \Vdash^{\mathcal{C}} (E \text{ in } Z) >_a (E \text{ in } S)$ and $\mathfrak{A}, i \Vdash^{\mathcal{C}} (E \text{ in } Z) >_b (Z \text{ in } N)$, then by definition there are **CS** models (r_a, ρ) and (r_b, ρ), such that $(r_a, \rho), i \Vdash (E \text{ in } Z) > (E \text{ in } S)$and $(r_b, \rho), i \Vdash (E \text{ in } Z) > (Z \text{ in } N)$. But there is no mutual refinement of r_a and r_b, i.e. there is no $c \in \mathcal{C}$ such that $r_c \in \mathcal{R}[r_a] \cap \mathcal{R}[r_b]$, for the simple reason that refinements are $>$-truth preserving, by *Proposition 4.5*. If there was such a mutual refinement r_c, then $(r_c, \rho), i \Vdash \{(E \text{ in } Z) >_a (E \text{ in } S), (E \text{ in } Z) >_b (Z \text{ in } N)\}$, which we have established in *Claim 4.5.2* to be impossible, since (r_c, ρ) is a **CS** model, by definition. □

Explanation: contextual information incompatibility (contextual orthogonality)

Although on the contextualized analysis there are now contexts *a* and *b* such that both premises $A >_a B$ and $A >_b C$ can be true at some possible world *i*, according to r_a and r_b respectively (in contrast with the case of C and its extensions and with the case of **CS** and its extensions), but there is no context *c* such that r_c is a mutual refinement of r_a and r_b. In other words, it is not possible to integrate the contextual information of contexts *a* and *b*, carried by ordering assignments of ordering frames r_a and r_b to any world i in a manner that corresponds

to some possible context *c* whose information would be carried by the ordering assignment of ordering frame r_c to world i.[170]

For the same reasons the following instance of Adjunction of Consequents goes through, where the implicit inconsistency is a little more obvious than in the 'Everest in NZ' example. Here the consequents of the premises form an inconsistent set, whereas in the other example the addition of the antecedent (and obvious enthymemes) to the consequent pair resulted in inconsistency.

Example 4.5.4: $A >_a B, A >_b {\sim}B \vDash^{\mathcal{C}}_{\mathbf{CS1+}} A >_c (B \wedge {\sim}B)$

4.4.5.5 Fine-tuning CS1+ and the system CS2+

This motivates the following reformulation of valid inference conditions, which do away with (i.e. block) cases when the ordering frames corresponding to contexts that make all the premises true fail to have a mutual refinement.

Definition 4.5.5: Let $\mathfrak{A}$ be a **CS**+ model. For all $\Sigma \cup \{A\} \subseteq For^{\mathcal{C}}_{>0}(>) \cup \overline{For^{\mathcal{C}}_{>}}$:

$\Sigma \vDash^{\mathcal{C}}_{\mathfrak{A}} A$ iff

(1) $\exists \mathfrak{F} \in \mathbf{CS}$, and

(2) If for all $i \in W$:

(i) $\mathfrak{A}, i \Vdash^{\mathcal{C}} \Sigma$, and

(ii) if $Ind(\{A\}) = \emptyset$, then $\mathfrak{F} \in \bigcap\{\mathcal{R}[r_b] : b \in Ind(\Sigma)\}$, and

if $|Ind(\{A\})| = 1$, then $\mathfrak{F} = r_a \in \bigcap\{\mathcal{R}[r_b] : b \in Ind(\Sigma)\}$ for $\{a\} = Ind(\{A\})$,

then $\mathfrak{A}, i \Vdash^{\mathcal{C}} A$.

Note that adding condition (1), which requires the existence of a mutual refinement of ordering frames that represent the context-indices over which the premises range will make $\Sigma \vDash^{\mathcal{C}}_{\mathfrak{A}} A$ false, for each **CS**+ model $\mathfrak{A}$, if no such refinement exists. Precisely what is required to invalidate *Adjunction of Consequents*, by paralleling our intuitions in the treatment of the

---

[170] Contrast example 4.5.1 with the following one, based on an example from Quine, where the contexts required to make the premises true need not be incompatible as is the case with ones that make the premises of 4.5.1 true.

(1) If Caesar had been in command, he would have used the atom bomb.

(2) If Caesar had been in command, he would have used catapults.

(3) Therefore, if Caesar had been in command, he would have used catapults and the atom bomb.

Both premises can be true on a single, albeit rather eccentric, context and the conclusion is also naturally true in that context. This doesn't change the fact that the inference form is invalid.

running counterexample.
(Observation: condition (1) resembles in its form the syntactic, *propositional variable sharing condition* for valid relevant conditionals, i.e. *Definition 1.7*.)

Definition 4.5.6: For all $\Sigma \cup \{A\} \subseteq For_{>0}^{\mathcal{C}}(>) \cup \overline{For_{>}^{\mathcal{C}}}$: write $\Sigma \vDash_{\mathbf{CS2+}}^{\mathcal{C}} A$ iff $\Sigma \vDash_{\mathfrak{A}}^{\mathcal{C}} A$ for all **CS**+ models $\mathfrak{A}$, where $\Sigma \vDash_{\mathfrak{A}}^{\mathcal{C}} A$ is as defined in 4.5.5. Also write $\vDash_{\mathbf{CS2+}}^{\mathcal{C}} A$ iff $\vDash_{\mathfrak{A}}^{\mathcal{C}} A$ for all **CS**+ models $\mathfrak{A}$. Call this logic (schema) **CS2**+.

What is paradigmatic about such inference forms is that they highlight what is really at play in contextualized validity when we explore limit cases, i.e. premises being true in radically different contexts. That is, we can have possibility expressing premises true for any contexts, but the inference is valid if the conclusion can always be true in a contextually meaningful way – one that is not independent of the contextual information by virtue of which the premises are true. If there is no mutual refinement of ordering frames representing context-indices over which the premises range, that means there is no single context on which all the premises are true, and consequently no contextually meaningful way of speaking of the conclusion following from those premises. Therefore, the inference is contextually invalid. It should be noted that the inference fails in limit cases as exemplified in 4.5.1 but may very well go through on some **CS**+ models (if not all) if the divergence of contexts over which the premises range isn't extreme.

It could be argued that such contextual incompatibility of premises – all true but on contexts that do not have a mutual refinement – should be treated in the manner that inconsistent sets of premises are treated, i.e. the conclusion should follow vacuously. Perhaps this needs some more thought, but examples such as 4.5.1 – which appear to be legitimate counterexamples to *Adjunction of Consequents* – seem to speak against such an approach. The inference is invalid, and it is only the contextualized account that gives the corresponding correct analysis, allowing for the premises to be jointly (and meaningfully) true.

Chapter 5

A non-vacuist account of counterpossibles

Everyone knows that dragons don't exist. But while this simplistic formulation may satisfy the layman, it does not suffice for the scientific mind. [...] Indeed, the banality of existence has been so amply demonstrated, there is no need for us to discuss it any further here. The brilliant Cerebron, attacking the problem analytically, discovered three distinct kinds of dragon: the mythical, the chimerical, and the purely hypothetical. They were all, one might say, non-existent, but each non-existed in an entirely different way.

Stanisław Lem, *The Cyberiad*, 1965.

5.0 Introduction

In this chapter I develop a non-vacuist account of counterpossibles, by building on the ordering semantics given for the system **CS** in chapter 4. That is, I modify **CS** in a manner that results in an analysis of counterpossibles that meets our intuitions, i.e. as sometimes being non-vacuously true and sometimes non-vacuously false. As we saw in chapter 2, one of the major drawbacks of Lewis' account of the counterfactual is that it evaluates all counterfactuals with impossible antecedents (*viz*. counterpossibles) as true, including intuitively false ones like:

(i) If Alice had squared the circle and Bob had doubled the cube, then Alice would be Bob.
(ii) If paraconsistent logic were correct, *ex contradictione quodlibet* would still be valid.

This stems from the fact that Lewis' (1973, 1981) analysis of counterfactuals is restricted to possible worlds, which results in all counterpossibles satisfying the truth conditions

vacuously.[171] The inadequacy of *vacuism*, as such analyses have come to be known in the context of discussions of counterpossibles, has already been identified and challenged by a number of authors.[172] I join this critical front, and drawing on existing proposals for non-vacuism, show that there is a sense in which we can preserve all of Lewis' analysis of mere counterfactuals, whilst avoiding the vacuous truth of counterpossibles, by admitting impossible worlds as worlds where the impossible is true.[173]

In §5.1 I present a family of logics and their ordering semantics, based on partial preorderings of worlds, i.e. systems based on ordering frames, much like the **CS** ordering frames of chapter 4, but where we allow our universal quantifier in the truth conditions for $>$ to range over impossible worlds, and where a new set of conditions on the ordering (comparative similarity) of worlds is introduced in order to accommodate for this domain extension. In section §5.2 I demonstrate that all the systems characterized in §5.1 meet the non-vacuity criterion, i.e. they are systems in which some counterpossibles *are* false. §5.3 is devoted to a critical discussion of the counterpart to Nolan's (1997) *Strangeness of Similarity Condition* (SIC), where I argue in SIC's favour based on the evaluation of the benefits and costs of its implementation. In §5.4 I discuss the matter of comparability of worlds on the extended account. Most of that discussion has the character of a reply to Weiss' (2017) objection to the general idea underlying similarity semantics, i.e. *comparability* of worlds, which he thinks to be a sufficiently fundamental hindrance to question the tenability of a similarity approach to analyzing counterpossibles. He sets his objection in terms of an alleged counterexample to an inference form that is valid on all similarity systems that satisfy comparability of worlds. My reply shows that the challenged inference is invalid on some weaker systems, proposed in §5.1 – namely, those satisfying a weaker ordering condition, whereby comparability is only lifted from impossible worlds. I also argue that an apparent lack of clarity – in the formulation of Weiss' counterexample – regarding the permissible extent of contextual shift between the reading (as true) of the different pertinent premises, could be used to invalidate more than he has intended. I close §5.4 by showing that the weaker systems invalidate *Adjunction of Consequents* even for mere counterfactuals.

[171] See definition 2.19 (4) and definition 4.2.5. (8).
[172] Nolan (1997), Mares (1997), Lander Laan (2004), Brogaard and Salerno (2008), Priest (2008), Bjerring (2014), Weiss (2017), Berto & Jago (2019, §12).
[173] The earliest use of the term 'non-vacuism' in this context appears in Brogaard and Salerno (2014).

5.1 Ordering semantics for counterpossibles

The formal language of the analysis is propositional modal logic with an additional dyadic modal connective $>$. That is, the definition of the formal language $\mathcal{L}$ and the set For of well-formed formulae is as defined in §4.1.

To simplify the discussion in this chapter we will work with a restricted class of ordering frames, where *all worlds are accessible*, i.e. recalling **CS** ordering frames of the previous chapter, in this chapter we let $S_i = W$ for each $i \in W$ for each ordering, so all talk of accessibility is set aside, while we focus on pertinent features of ordering frames that are most relevant to counterpossibles.[174] I will be considering a weaker foundational ordering condition than (CS1) of **CS** ordering frames as the basis for modelling comparative similarity over the extended domain (impossible worlds), with optional conditions that allow strengthening it for possible worlds. That is, the additional conditions allow some worlds to be incomparable in terms of their similarity to the actual world (or any world of evaluation) – a property of comparative similarity ruled out by Lewis (1971, 1973, 1981) and Stalnaker (1968, 1970).

Definition 5.1: An *ordering frame* is a triple $(W, N, \{\lesssim_i : i \in N\})$, where W is a nonempty set, $\emptyset \neq N \subseteq W$ and $\lesssim_i \subseteq W \times W$ satisfies the following conditions, for each $i \in W$:

(CS1.1) $\lesssim_i$ is a preorder on W

On the intended interpretation, elements of N are possible (or normal) worlds, $W \setminus N$ are impossible (or non-normal) worlds, and $\lesssim_i$ is regarded as the ordering of worlds in respect of their comparative similarity to i, with the following intended meaning:

$j \lesssim_i k$: j is at least as similar to i as k is.
$j <_i k$: j is more similar to i than k is.
$j \sim_i k$: j and k are equally similar to i.

Only possible worlds are given a comparative similarity assignment (comparative similarity neighbourhood), since on this picture truth at impossible worlds will be independent of any similarity considerations.

[174] In this section and the next I simplify the formulation by borrowing some layout and presentation features of the model theory from Sillari (2008, §2.3), which draws on Hintikka and Rantala's work, Priest (2008, §9.4.7), Weiss (2017) and Berto & Jago (2019).

Definition 5.1.1: Denote the *class of ordering frames from Definition 5.1* with **CS***.

I'd like to highlight a property of sets that play a key role in our semantics, one which allows me to introduce a short-hand notation that will aid most formal arguments in this chapter.[175]

Definition 5.2: Given a preordered set $(S, \precsim)$. Call $I \subseteq S$ an *ideal* in $(S, \precsim)$ iff

(i) $I \neq \emptyset$

(ii) I is a lower set: $(\forall x \in I)(\forall y \in S)(y \precsim x \longrightarrow y \in I)$.

(iii) I is a directed set: $(\forall x, y \in I)(\exists z \in I)(x \precsim z \wedge y \precsim z)$.

Definition 5.2.1: Let $(S, \precsim)$ be a preordered set and $x \in S$. Define $\downarrow x_{(S,\precsim)} := \{y \in S: y \precsim x\}$. When there is no ambiguity regarding the preordered set in question, I'll omit the subscript.

Proposition 5.1: Let $(S, \precsim)$ be a preordered set such that $x \in S$. $\downarrow x$ is an ideal in $(S, \precsim)$.
Proof : Condition (i) is immediate, since $x \in \downarrow x$, as is (ii) from the definition of $\downarrow x$, and (iii) follows from the fact that $x \in \downarrow x$ and $y \precsim x$ for all $y \in \downarrow x$. □

The formulation of the model theory is relatively common, and can be traced back as far as Kripke's semantics for C.I. Lewis' systems that are weaker than S4, which we looked at in §1.3, with the additional feature borrowed from Rantala's models, which we looked at in §1.3.4, of assigning arbitrary values to formulae at non-normal worlds – a method that has been widely applied in impossible world semantics for doxastic and epistemic logics that model non-ideal agents, i.e. that avoid logical omnidoxasticity and omniscience.[176]

Definition 5.3: A model *based on ordering frames* is the quadruple $\mathfrak{A} = (W, N, \{\precsim_i: i \in N\}, \rho)$ where $(W, N, \{\precsim_i: i \in N\})$ is an *ordering frame* and $\rho = \{\rho_i: i \in W\}$ is defined as follows:

(1) For $i \in N$: $\rho_i \subseteq PV \times \{0,1\}$ is a relation satisfying the following constraints:

(i) For no $p \in PV$ and $i \in W$, both $p\rho_i 0$ and $p\rho_i 1$ (exclusion)

(ii) For all $p \in PV$ and $i \in W$, either $p\rho_i 0$ or $p\rho_i 1$ (exhaustion)

(2) For $i \in W \setminus N$: $\rho_i \subseteq For \times \{0,1\}$.

[175] Ideals also appear in the truth conditions for $A > B$ in CS models, defined in chapter 4.
[176] E.g. see Sillari (2008, §2.3) who draws on the earlier work of Hintikka and Rantala.

That is, (2) tells us that truth values are related (assigned) directly to any formula at each non-normal world. Also note that the manner in which ρ_i's are restricted in (1) effectively renders them as functions $\rho_i: PV \longrightarrow \{0,1\}$ for each $i \in N$.

Definition 5.3.1: Truth in a model is defined in terms of the relation $\Vdash \subseteq N \times For$, defined as follows: given a model $(W, N, \{\lesssim_i : i \in N\}, \rho)$ and any $i \in W \setminus N$, and $A \in For$:

(1) $i \Vdash A$ iff $A\rho_i 1$

And for any $i \in N$, $p \in PV$, and $A, B \in For$:

(2) $i \Vdash p$ iff $p\rho_i 1$
(3) $i \Vdash {\sim}A$ iff not $i \Vdash A$
(4) $i \Vdash A \wedge B$ iff $i \Vdash A$ and $i \Vdash B$
(5) $i \Vdash A \vee B$ iff $i \Vdash A$ or $i \Vdash B$
(6) $i \Vdash A \supset B$ iff $i \Vdash {\sim}A$ or $i \Vdash B$
(7) $i \Vdash \Box A$ iff $\forall j \in N: j \Vdash A$.
(8) $i \Vdash \Diamond A$ iff $\exists j \in N: j \Vdash A$.
(9) $i \Vdash A > B$ iff $\exists k \in W: k \Vdash A$ and $\forall j \in W\big(j \lesssim_i k \longrightarrow (j \Vdash A \longrightarrow j \Vdash B)\big)$

The intended meaning of $i \Vdash A$ is 'A is true at i'. For each $i \in W \setminus N$, $\rho_i \subseteq For \times \{0,1\}$ is a relation, as specified in (1), between *any* formula and {0,1}. That is, truth conditions for complex formulae are not defined recursively, but related by ρ_i to complex formulae directly. This allows for the inclusion of *closed worlds*, where the laws of logic are different (e.g. the worlds may be closed under paraconsistent or paracomplete consequence), and *open worlds*, where even extensional formulas fail to conform to any rules of compositionality.

Definition 5.4: It will also be convenient to define $[A]^{\mathfrak{A}} := \{i \in W: \mathfrak{A}, i \Vdash A\}$ for any model $\mathfrak{A}$ with domain W. The superscript will be omitted when its absence will not lead to ambiguity.

Notation: with the help of *definition 5.4* let us formulate the second conjunct of (9) of *definition 5.3.1* more succinctly – the main reason for defining the notation $\downarrow.k$, in the first place.

(9') $i \Vdash A > B$ iff $\exists k \in W: k \Vdash A$ and $\downarrow k_i \cap [A] \subseteq [B]$.

Where $\downarrow k_i$ is just shorthand for $\downarrow k_{(W, \lesssim_i)}$. In contexts where the subscript is constant, I'll omit it altogether and just write $\downarrow k$ for brevity.

Logical truth is defined as truth at all possible worlds in each model, and valid inference is truth preservation at all possible worlds in each model. This follows an approach that can be traced at least back to Kripke's semantics for non-normal modal logics (see §1.2) and is a common approach to defining validity and valid inference in semantics that include non-normal or impossible worlds. The motivation for this definition of logical truth and validity is justified if we characterize impossible worlds to be those where the laws of logic are different or where the laws of logic fail. Then when we define validity and valid inference, i.e. the laws and rules of logic, we should not consider worlds where the laws of logic are different or where they fail.[177]

Definition 5.5: Let $\vDash_{\mathbf{CS}^*} \subseteq \wp(For) \times For$. Write $\Sigma \vDash_{\mathbf{CS}^*} A$ if and only if for all models $(W, N, \{\lesssim_i : i \in N\}, \rho)$, and all $i \in N$, if $i \Vdash B$ for all $B \in \Sigma$, then $i \Vdash A$. We say an inference from Σ to A is valid iff $\Sigma \vDash_{\mathbf{CS}^*} A$. That is, valid inference is defined as truth preservation at all possible worlds in all $\mathbf{CS}^*$ models. A formula $A \in For$ is said to be valid iff $\emptyset \vDash_{\mathbf{CS}^*} A$. Call this logic $\mathbf{CS}^*$.

Note that since the truth conditions for $\Box$ and $\Diamond$ formulae are defined in terms of unrestricted quantification over possible worlds, the above validity conditions give the modal logic **S5** for the basic modal language.

There are a number of additional, well-motivated conditions that one could impose on ordering frames, thereby generating a whole family of logics.

Definition 5.6: Given a **CS*** ordering frame $\mathfrak{F} = (W, N, \{\lesssim_i : i \in N\})$ define the following conditions on $\mathfrak{F}$, for all $i \in W$:

(WC)	i is $\lesssim_i$-minimal: $\forall j \in W(i \lesssim_i j)$.
(SC)	i is $<_i$-minimal: $\forall j \in W(j \neq i \longrightarrow i <_i j)$.
(T1)	$\lesssim_i$ is total over N: $\forall j, k \in N(j \lesssim_i k \vee k \lesssim_i j)$.
(T2)	$\lesssim_i$ is total: $\forall j, k \in W(j \lesssim_i k \vee k \lesssim_i j)$.
(SI1)	Possible worlds are $\lesssim_i$-minimal: $\forall j, k \in W(k \notin N \longrightarrow j \lesssim_i k)$.
(SI2)	Possible worlds are $<_i$-minimal: $\forall j, k \in W\big((j \in N \wedge k \notin N) \longrightarrow j <_i k\big)$.

[177] Berto & Jago (2019, §4.2)

If a **CS*** frame $\mathfrak{F}$ satisfies some condition (C), we will say '$\mathfrak{F}$ satisfies (C)'.

Definition 5.7: Let (C) be a condition predicable of an ordering frame (e.g. like the conditions in definition 5.6). Denote the restricted class of frames $\{\mathfrak{F} \in \mathbf{CS}^*: \mathfrak{F} \text{ satisfies (C)}\}$ with $\mathbf{CS}^*_{(C)}$.

Definition 5.5.1: Let $\vDash_{\mathbf{CS}^*+(C)}$ be defined as follows: let $(W, N, \{\precsim_i : i \in N\}) \in \mathbf{CS}^*_{(C)}$, and write $\Sigma \vDash_{\mathbf{CS}^*+(C)} A$ iff for all models $(W, N, \{\precsim_i : i \in N\}, \rho)$ and all $i \in N$, if $i \Vdash B$ for all $B \in \Sigma$, then $i \Vdash A$. We say an inference from Σ to A is valid iff $\Sigma \vDash_{\mathbf{CS}^*+(C)} A$. That is, valid inference is defined as truth preservation at all possible worlds in all $\mathbf{CS}^* + (C)$ models. A formula $A \in For$ is said to be valid iff $\emptyset \vDash_{\mathbf{CS}^*+(C)} A$. Call this logic $\mathbf{CS}^* + (C)$.

Note that the basic (CS1) condition of total preorderhood over the entire domain, used in the previous chapter is equivalent to the conjunction of the weaker condition of mere preorderhood and unrestricted totality, i.e. (CS1.1)+(T2). Conditions (SI2) and (SI1) correspond to Nolan's (1997, p.566) conditions, i.e. *Strangeness of Impossibility* condition (SIC) and Lesser *Strangeness of Impossibility* condition (LSIC), respectively. (SI1) is the weaker of the two, as it only demands that no impossible world is more similar to the world of evaluation than some possible world, whereas (SI2) stipulates that all possible worlds are more similar to the world of evaluation than any impossible world. Aside from its intuitive appeal (SI2) has also important formal advantages, and it is not entirely free of criticism, all of which I'll address in §5.3.

As I will argue in §5.4 there may be good reasons to think that the notion of comparative similarity modelled by total preorders may be too strong when it comes to impossible worlds, so we could weaken the orderings from being total preorders of the entire domain (CS1.1)+(T1), which are jointly equivalent to (CS1), to only being totally preordered over possible worlds (CS1.1)+(T2) thereby allowing incomparabilities between impossible worlds. The intuition here is that impossible worlds are so strange that it would seem a little strong to demand that even in relevant respects their conceptual impossibility (logical, mathematical, metaphysical, etc.) should always be comparable by $\precsim_i$. That is, for any two impossible

worlds, it need not always be correct to say that one is more/less/equally similar than the other to the actual world (or any world of valuation), even in relevant respects.

The primary aim of the discussion in this chapter is to focus on weighing up the pros and cons of the following extensions (with convenient denotations indicated) of $\mathbf{CS}^*$.

$\mathbf{CS}_1^*$: $\mathbf{CS}^* + (\mathrm{WC})$
$\mathbf{CS}_2^*$: $\mathbf{CS}^* + (\mathrm{T1}) + (\mathrm{WC})$
$\mathbf{CS}_3^*$: $\mathbf{CS}^* + (\mathrm{T1}) + (\mathrm{WC}) + (\mathrm{SI2})$
$\mathbf{CS}_4^*$: $\mathbf{CS}^* + (\mathrm{T2}) + (\mathrm{WC}) + (\mathrm{SI2})$

Note that $\vDash_{\mathbf{CS}_1^*} \subseteq \vDash_{\mathbf{CS}_2^*} \subseteq \vDash_{\mathbf{CS}_3^*} \subseteq \vDash_{\mathbf{CS}_4^*}$ by definition. The system $\mathbf{CS}_2^*$ is a lot like **CS**, introduced in the previous chapter, with the only difference that $\mathbf{CS}_2^*$ models admit impossible worlds with a weaker condition modelling comparative similarity between them. Its extension $\mathbf{CS}_3^*$ adds the Strangeness of Impossibility Condition (SI1) – which offers a number of advantages (discussed in §5.3) – and as I argue in §5.4, may be a better option than $\mathbf{CS}_4^*$. In other words, I will argue that $\mathbf{CS}_3^*$ is the optimal system, relative to the ones considered here.

5.2 Adequacy for non-vacuism of the weakest CS* systems

It's easy to check that $\mathbf{CS}^*$ validates the law of identity:

(5.1) $\vDash A > A$

And the addition of (WC) to $\mathbf{CS}^*$ validates modus ponens and modus tollens:

(5.2) $A, A > B \vDash B$

(5.3) $\sim B, A > B \vDash \sim A$

CS* and its extensions are non-vacuist, which is the first indication of their adequacy. That is, both of the following no longer hold in $\mathbf{CS}_1^*$ and its extensions:

(5.4) $\sim\Diamond A \vDash A > B$ e.g. $\sim\Diamond(\sim p \wedge p) \nvDash_{\mathbf{CS}^*} (\sim p \wedge p) > q$

Which as a consequence results in the invalidation of:

(5.5) $\Box(A \supset B) \vDash A > B$ e.g. $\big((\sim p \wedge p) \supset q\big) \nvDash_{\mathbf{CS}^*} (\sim p \wedge p) > q$

Proposition 5.2: $\nvDash_{\mathbf{CS}^*} (\sim p \wedge p) > q$

Proof : Let $\mathfrak{A} = (W, N, \{\lesssim_i : i \in N\}, \rho)$, be a **CS*** model such that $W = \{i, j\}$, $N = \{i\}$, $i \lesssim_i j$, and $(\sim p \wedge p, 1) \in \rho_j$ and $(q, 1) \notin \rho_j$. So, $i \nVdash (\sim p \wedge p) > q$, since $j \Vdash \sim p \wedge p$ and $j \lesssim_i j$, but $j \nVdash q$. □

Corollary 5.2.1: $\Box\big((\sim p \wedge p) \supset q\big) \nvDash_{\mathbf{CS}^*} (\sim p \wedge p) > q$

Proof : Follows immediately from proposition 5.2, since $\vDash_{\mathbf{CS}^*} (\sim p \wedge p) \supset q$. □

5.3 Strangeness of Impossibility Condition

When evaluating at a possible world the truth of a counterfactual whose antecedent doesn't express an impossibility, impossible worlds are irrelevant, much like worlds where kangaroos walk upright using crutches are irrelevant in evaluating the counterfactual 'If kangaroos had no tails, they would topple over'. So, the condition (SI2) is very much Lewisean in spirit.[178] There's an obvious parallel between the centering conditions (SC) and (WC) and strangeness of impossibility conditions (SI1) and (SI2). Just as (SC) stipulates that the actual world (or any world of evaluation) is more similar to itself than all other worlds, (SI2) stipulates that all possible worlds are more similar to the actual world (or any world of evaluation) than any impossible world is. Both (WC) and (SI1) weaken those conditions by allowing ties in comparative similarity between the world of evaluation and other possible worlds and by allowing ties in comparative similarity between possible worlds and impossible worlds, respectively.[179] Perhaps unsurprisingly, to that analogue in comparative similarity restrictions there corresponds a pair of characteristic inference forms that hinge on them.[180]

(5.6) $A, B \vDash A > B$

(5.7) $\Box A, \Box B \vDash A > B$

That is, (5.6) is invalidated on all $\mathbf{CS}^*$systems that do not satisfy (SC), so on the current proposal that means all systems $\mathbf{CS}_1^*$ through $\mathbf{CS}_4^*$. Also, systems that satisfy (SI2), validate (5.7), which I prove shortly in *Proposition 5.3*, so on the current proposal only systems $\mathbf{CS}_3^*$ and $\mathbf{CS}_4^*$ validate it. In §5.3.2.2 I will address what looks like a counterexample to (5.7) given by Weiss (2017), and which arms his objection to (SI2).

5.3.1 Benefits

The main appeal of (SI2) is that it allows us to effectively preserve all of Lewis' analysis of mere counterfactuals, whilst correcting the analysis of counterpossibles. Adding the stronger strangeness of impossibility condition (SI2) validates (5.7) and:

(5.8) $\Diamond A, A > B \vDash \Diamond B$

---

[178] Mares (1997) and Jago (2014) also endorse (SIC).

[179] We could think of (SI1) as "weak centering on N" and (SI2) as "strict centering on N".

[180] Weiss (2017, §2.2) also makes an analogous observation.

Proposition 5.3: $\Box A, \Box B \vDash A > B$

Proof : Let $\mathfrak{A} = (W, N, \{\lesssim_i : i \in N\}, \rho)$ be a **CS*** model that satisfies (SI2), and let $i \Vdash \{\Box A, \Box B\}$ for arbitrary $i \in N$. Then it follows that there is a $j \in N \subseteq [A]$ and $\downarrow j \cap [A] \subseteq [B]$ since $\downarrow j \cap [A] \subseteq N$ by hypothesis and (SI2), and $N \subseteq [B]$ by hypothesis. So, $i \Vdash A > B$, as required. □

Proposition 5.4: $\Diamond A, A > B \vDash \Diamond B$

Proof : Let $\mathfrak{A} = (W, N, \{\lesssim_i : i \in N\}, \rho)$ be a **CS*** model that satisfies (SI2), and assume $i \Vdash \Diamond A$ and $i \Vdash A > B$ for arbitrary $i \in N$. Now, since $N \cap [A] \neq \emptyset$, then $i \Vdash A > B$ and (SI2) imply that there exists a world $j \in N \cap [A]$ such that $\downarrow j \cap [A] \subseteq N$ and $\downarrow j \cap [A] \subseteq [B]$. Hence, $j \Vdash B$, which implies that $i \Vdash \Diamond B$, as required. □

As a matter of fact (SI2) does a lot more. It allows us to salvage most of the **CS** valid inferences, as long as the antecedents of the modal conditional $>$ are restricted to expressing possible propositions. This can be done by adding suppressed premises in the form $\Diamond A$ to the premise set, for every $A > B$ appearing anywhere in the inference, which would ensure that when evaluating a conditional $>$ we would never "look beyond" possible worlds. That is, given a **CS** valid inference $\Sigma \vDash_{\mathbf{CS}} A$, we can preserve its validity on **CS*** systems if we add $\Diamond B$ to the premise set for each $B > C \in Sub(A) \cup \{D \in Sub(E) : E \in \Sigma\}$.[181] In order words, with (SI2) in place we preserve all of Lewis' analysis of mere counterfactuals (i.e. possible-antecedent part of **CS**), whilst correcting the analysis of counterpossibles. That is, by introducing impossible worlds we lose nothing of the original analysis, and we gain by amending its drawbacks.

To be sure (SI2) is not entirely unobjectionable. Nolan (1997) having introduced (SI2) not so much as a logical principle, but a tentative heuristic, explores a few insightful examples that could be said to violate it. Berto and Jago (2019) reply to those examples by defending (SI2). I will not reiterate that exchange here but instead focus on a couple of other objections, in the next section.

5.3.2 Criticisms

5.3.2.1 The problem of the trivial world

A number of authors have observed an internal tension between two provisional principles that have gathered wide acceptance by those working with similarity semantics for

[181] In agreement with (Berto & Jago, 2019).

counterpossibles based on classical logic.[182]

Definition 5.8: Given any **CS*** model call a world $w \in W$ closed under L-consequence if and only if $\Sigma \vDash_L A$ and $w \Vdash \Sigma$ implies $w \Vdash A$, where L is some logic.
We could then say that a world *w* is governed by logic L iff *w* is L-closed.

The first of the aforementioned principles is (SI2) and the second is the suggestion that the world λ, where everything is true, i.e. a world $\lambda \in W \backslash N$ such that $(\forall A \in For)[\lambda \Vdash A]$, also referred to in the literature as the *trivial world* (aka *explosion world*, *absurd world*) should be relegated to be among the most dissimilar impossible worlds (or at least no less dissimilar than any other impossible world), which I'll denote with (ST) for *strangeness of the trivial world*.[183] The aforementioned tension stems from the fact that some of the reasons that speak in support of (SI2) simultaneously – it could be argued – speak against (ST). Namely, closure under classical consequence is part of the justification for (SI2), but the trivial world is also closed under classical consequence – in particular, unlike other LNC-violating worlds, it satisfies ECQ.[184] So, if we were to attribute principal weight (although we don't) to logical closure in determining the similarity of worlds – which is part of the justification for (SI2) – then that would not only speak against (ST), but strongly in favour of the trivial world being the most similar impossible world. But this would result in trivializing the analysis once more, since the consequent of any counterpossible would be true at the closest antecedent-admitting world, namely λ.[185] So, closure under classical consequence can't be the only criterion for determining the comparative similarity of impossible worlds. On the current proposal the formalism to avoid such trivialization is in place, since the trivial world need not be present in every model, and even when it is present, ties between impossible worlds are allowed. The obvious justification for a variation of such orderings is that we don't always attribute priority to logical closure in determining the similarity of impossible worlds. That is,

[182] (Sendłak, 2016, 2017), (Weiss, 2017).
[183] See (Stalnaker 1968, p.103), (Nolan 1997,p.544), (Berto 2013), and (Brogaard & Salerno, 2014, p.652). Stalnaker (1968) denotes the absurd world with λ. Note that on the relational semantics approach presently chosen for **CS*** models, there is a whole class $\{w \in W \setminus N: (\forall A \in For)[w \Vdash A]\}$ of impossible worlds that are closed under classical consequence, since if A is a truth value glut at some world i, i.e. $A\rho_i 1$ and $A\rho_i 0$, then $\Vdash A$, by definition.
[184] LNC stands for the *law of non-contradiction*, i.e. $\vDash \sim(A \wedge \sim A)$ for any $A \in For$ (a proposition and its negation can't both be true), and ECQ stands for *ex contradictione quodlibet*, i.e. $A, \sim A \vDash B$ for any $A, B \in For$ (anything follows from a contradiction).
[185] It would be equivalent to the vacuous analysis given by Lewis, but resemble in its formalism Stalnaker's approach, who stipulated the absurd world λ to account for counterpossibles.

it is not the case that in each context in which we entertain an impossible scenario, we always import logical closure as the information relevant to the evaluation of the counterpossible. In the next couple of paragraphs, I argue that closure under classical consequence is insufficient to justify deeming λ the closest impossible world.

There are many reasons that speak against treating λ as the closest impossible world. First, although it is closed under classical consequence, it also is "maximally inconsistent", in the sense that $\lambda \Vdash \{A, {\sim}A\}$ for all $A \in For$. Surely the degree to which a world is LNC-violating should factor in to its similarity. Another way of looking at this is to consider what *doesn't hold* at a world – surely this is not entirely irrelevant and should also feature as a similarity parameter. When we take into account what fails to be true at any given world, then λ departs in the greatest possible way from any possible world *w*, because $\lambda \Vdash A$ for any $w \nVdash A$.[186]

Next, λ isn't only closed under *classical* consequence – it is closed under *any* truth preserving consequence. So, it's unclear why λ should be strictly closer to (classically) possible worlds than other LNC-violating worlds that are closed under any other truth preserving consequence. Deeming such worlds to be at least as similar to any possible world as λ is, seems like a perfectly natural comparative similarity ordering.

Therefore, it's not entirely clear that logical closure justification for (SI2) speaks in favour of λ being the closest impossible world. (SI2) is a statement regarding the difference in similarity between *possible* and *impossible* worlds. The fact that λ happens to be closed under classical consequence (given that it's closed under all truth preserving consequences) at best speaks in favour of it having a similarity advantage over open worlds, i.e. worlds a lot like λ, where things hold for no reason, and which have been designed to violate all closures. That is, the claim that in all contexts λ should be the deemed as the most similar impossible world appears to be false.

5.3.2.2 Other objections

Weiss (2017) gives an example that questions the validity of (5.7), which is effectively an objection to (SI2), which is sufficient to validate (5.7). The counterexample goes as follows:

[186] This does become highly relevant on some informational interpretations of states of affairs (on some ersatz views of worlds), where the distinction – absent in classical possible worlds – between *negative* information and *absence* of information is key, e.g. see Mares (1997, §3).

let the first premise be 'there either is a counterexample to LEM[187] or there is no counterexample to LEM' and let the second premise be 'there is no counterexample to LEM'. Both are necessarily true, but the conclusion 'if it were the case that there either is a counterexample to LEM or there isn't a counterexample to LEM, then there would be no counterexample to LEM' doesn't seem to follow.

I've hinted earlier at parallels between (SI2) and (SC), which become salient in the similarities between (5.6) and (5.7). Apparently both inferences also raise analogous concerns – just as in the case of (5.6) we agree that the truth of the counterfactual doesn't depend on the mere (coincidental) truth of the antecedent and consequent (and indeed many counterexamples support that judgement), so in the case of (5.7) it appears that the truth of the counterfactual doesn't depend on the mere (necessary) truth of the antecedent and consequent, and the above example appears to support that judgement. In fact, the example given by Weiss can be viewed as a variation of *Hájek's* example aimed at challenging (5.6), which we have looked at in §2.2.7.[188]

Likewise, whereas abandoning (SC) in favor for (WC) can be motivated by an interpretation of similarity as *similarity in relevant respects* (where other worlds may be equally *similar in relevant respects* to the world of evaluation as it is to itself), it seems that an analogous motivation could speak in favor of (SI1) – which would suffice to invalidate (5.7) and which the offered formalism allows – but at the risk losing much of the mere counterfactual analysis (i.e. the possible-antecedent part of **CS**).

Note that (5.7) doesn't contain any *counterpossibles*, so it's not exactly a problem of the current proposal, but of the original analysis due to Lewis which validates it. The current proposal intends to heal the vacuous analysis of counterpossibles of Lewis' original account and employing (SI2) has proven sufficient for meeting that challenge, whilst offering a way to preserve most of the original analysis of mere counterfactuals. The matter of dealing with general relevance-failure issues, which (5.7) is a symptom of – although a matter certainly worth addressing – is sufficiently independent from the intended task of this chapter to be set aside for another time.

---

[187] LEM stands for law of excluded middle, i.e. $\vDash A \vee {\sim}A$ for any $A \in For$.

[188] See (§2.2.7, p.73). In the coin scenario we have a suppressed premise, which is an instance of LEM regarding all physical possible outcomes of the coin toss, and the second premise pertains to a fact, i.e. the coin having landed heads. In the example given by Weiss, the first premise is an instance of LEM regarding the existence of counterexamples of LEM, and the second premise is a statement of "logical fact" (we're assuming classical logic). Both conclusions seem wrong due to the ampliative character of the consequent.

5.4 The question of comparability of impossible worlds

Although its explicit form varies, depending on the particular semantic apparatus – be it sphere systems, selection functions, or ordering frames – comparability is the fundamental feature of similarity accounts of counterfactuals. Stalnaker (1968) and Lewis (1973, 1981), both agree that the orderings of worlds fitting the analysis of counterfactuals admit no incomparabilities – a condition that has been shown not to be necessary in general, but its appeal on a comparative similarity interpretation of orderings has some intuitive force.[189] I will not join that debate here, which is restricted to possible world semantics, but focus on reasons for lifting comparability from impossible worlds.

Such a move is partly motivated by a rebuttal to a general objection to comparative similarity semantics for counterfactuals, and other reasons that align with more general features of non-vacuism. I begin the section with a critical analysis and a reply to an objection by Weiss (2017) to the similarity account that takes aim at *comparability*, which lies at the heart of Lewis-Stalnaker comparative similarity semantics for counterfactuals.[190] I conclude the section by highlighting an additional aspect of CS* systems that proves beneficial to the analysis of mere counterfactuals, and which is gained from weakening the comparability conditions. That is, such systems are weak enough to correctly invalidate inference forms, which are nevertheless formally valid on a number of popular accounts of conditional logics, despite the existence of intuitive counterexamples.

5.4.1 Weiss' objection

First, I'll outline Weiss (2017) objection and show that (5.10) is valid on the strongest system $\mathbf{CS}^{*}_{4}$, characterized by ordering frames that satisfy the stronger totality condition (T2) whereby all worlds are totally preordered. Then I'll give a detailed account and critical analysis of Weiss' alleged counterexample to (5.10), and finally I'll show that (5.10) is invalidated on systems where (T2) is replaced by (T1).

Weiss (2017) objects to the inference rule (5.10), which holds for all CS logics characterized by ordering frames based on preorderings that are *total*.

[189] Notably Pollock (1976) and Kratzer (1981) effectively argue in favour of what would correspond to partial orderings on ordering semantics. For a good discussion of the various approaches see Lewis (1981, §3-5).

[190] Comparability is the basic assumption about comparative similarity of worlds that states: any two worlds *x* and *y* are comparable to each other in terms of their similarity relative to the world of evaluation *z*. *Totality* of preorders captures comparability for ordering frames, and *nesting* captures this for systems of spheres.

(5.10) $A > B, B > A \vDash (A > C) \equiv (B > C)$.[191]

To be exact, the objection is actually addressed to systems of spheres candidate semantics for a non-vacuist account of counterpossibles, and Weiss correctly identifies the *nesting condition* (see §2.2.3), fundamental for those systems, as responsible for the rule's validity.[192] Indeed (5.10) is characteristic of systems of spheres that are *nested*, and therefore all systems of spheres as defined by Lewis (1973). I'll mirror the discussion in terms of CS* models, noting that (5.10) is characteristic of ordering frames based on preorderings that are *total*, and therefore all CS models.[193] That is, comparability takes the form of *nesting* on systems of spheres and the form of *totality* on ordering frames based on preorders.

The objection is set up via what Weiss takes to be a counterexample to (5.10), formulated in terms of counterpossibles, and – the argument goes – because all systems of spheres satisfy nesting, a successful counterexample to (5.10) amounts a counterexample to sphere semantics in general. This objection extends to all other formulations that encode the intuitions about comparative similarity in terms of conditions corresponding to *comparability* – the basic intuition regarding comparative similarity of worlds. Therefore, in particular it is also an objection to ordering semantics based on *total* preorderings. However, Weiss' conclusion is too strong. Surely the alleged counterexample alone, even if correct, doesn't justify abandoning comparability altogether, but rather at most justifies lifting the nesting condition for spheres containing impossible worlds, or correspondingly in ordering semantics, lifting the totality condition from impossible worlds. And such a much weaker conclusion is not as damaging to similarity semantics.[194]

Proposition 5.2: $A > B, B > A \vDash_{\mathbf{CS}^*_4} (A > C) \equiv (B > C)$

[191] The axiomatic counterpart of (5.10) is CSO: $[(A > B) \wedge (B > A)] \supset [(A > C) \equiv (B > C)]$, see Nute (1980, §3.1).

[192] Weiss' (2017) sphere models for non-vacuism are based on sphere models equivalent to **S** models (see chapter 2), whose domains are extended to include non-normal worlds and the truth conditions at non-normal worlds are extended much in the same manner as I have modified **CS** models to yield **CS*** models (definition 5.4 and 5.4).

[193] *Nesting* and *totality* are each other's counterparts on **S** frames and **CS** frames, respectively. For a formal proof of that correspondence see lemmas *A*.1.0.1 and *A*.1.0.2 in the Appendix.

[194] An axiom, characteristic of all Lewis-Stalnaker logics of counterfactuals and closely related to (5.10), has been objected to before by Gabbay (1972) more generally, i.e. even in cases where the antecedent of the counterfactual doesn't express an impossibility. Gabbay objects to $\big((A > B) \wedge (B > A) \wedge (A > C)\big) \supset (B > C)$, by providing an insightful counterexample to what he believes as illustrating some relevance violating features of that inference. Note that its failure implies the failure of (5.10).

Proof : First, I'll prove that if $i \Vdash \{A > B, B > A\}$ for any $\mathbf{CS}_4^*$ model $(W, N, \{\lesssim_i : i \in N\}, \rho)$, $i \in W$, and $A, B \in For$, then $\exists k \in [A] \cap [B]$ such that $\downarrow k \cap [A] = \downarrow k \cap [B]$. Assuming $i \Vdash A > B$ implies $\exists k \in [A]$ such that $\downarrow k \cap [A] \subseteq [B]$, and $i \Vdash B > A$ implies $\exists k' \in [B]$ such that $\downarrow k' \cap [B] \subseteq [A]$. Note that in both cases $k, k' \in [A] \cap [B]$. Either $k \lesssim_i k'$ or $k' \lesssim_i k$, by (T2). Suppose $k \lesssim_i k'$. Hence $\downarrow k \subseteq \downarrow k'$. Now, $\downarrow k \cap [A] \subseteq [B]$ implies $\downarrow k \cap [B] \subseteq \downarrow k' \cap [B] \subseteq [A]$. Hence, $\downarrow k \cap [B] \subseteq [A]$. Hence finally, $k \in [A] \cap [B]$ and $\downarrow k \cap [A] = \downarrow k \cap [B]$. A similar argument shows that $\downarrow k' \cap [A] = \downarrow k' \cap [B]$ when $k' \lesssim_i k$. Let us denote such a world, which is guaranteed by $i \Vdash \{A > B, B > A\}$ with k^*. Now we will show that $i \Vdash A > C$ implies $i \Vdash B > C$, for any $C \in For$. Assuming $i \Vdash A > C$ implies $\exists k \in [A]$ such that $\downarrow k \cap [A] \subseteq [C]$. Next, by totality, either $k^* \lesssim_i k$ or $k \lesssim_i k^*$. Now, suppose $k^* \lesssim_i k$. Hence $\downarrow k^* \subseteq \downarrow k$, and we note that $\downarrow k^* \cap [B] = \downarrow k^* \cap [A] \subseteq \downarrow k \cap [A] \subseteq [C]$. Hence, $i \Vdash B > C$. Next, suppose $k \lesssim_i k^*$. Therefore $\downarrow k \subseteq \downarrow k^*$, which implies $\downarrow k \cap \downarrow k^* \cap [A] = \downarrow k \cap \downarrow k^* \cap [B]$. In conjunction with the hypothesis this implies $\downarrow k \cap [B] = \downarrow k \cap [A] \subseteq [C]$. Hence, $i \Vdash B > C$. The proof in the other direction is similar. So, $i \Vdash A > C$ iff $i \Vdash B > C$, as required. □

Now I'll focus on the alleged counterexample itself given by Weiss (2017), which arms the aforementioned general objection to most similarity accounts of counterfactuals and counterpossibles. Weiss (2017) formulates it as a variation on Williamson's (Hempel Lectures 2006, and Williamson 2007) objection to a non-vacuist account of counterpossibles, presented and discussed earlier in Brogaard and Salerno (2013, pp.649-50). I'll argue that the context-shift resulting from allowing (as true) certain premises opens the door to formulating other counterexamples that undermine inferences that are valid on all systems Weiss (2017) endorses as alternatives to similarity accounts to counterpossible analysis. But those premises are required for the counterexample to work.

The argument against (5.10) goes as follows:

> Fred asks George what 5+7 is, and George mistakenly responds 13. Fred snidely remarks, "if 5 + 7 were 13, you would have answered correctly." This is true. What else might be the case if 5 + 7 = 13? Plausibly, 5 + 6 = 12. Conversely, if 5 + 6 = 12, it would seem reasonable to expect that 5+7 = 13. From [5.10] and the truth of Fred's initial remark, we can infer "if 5+6=12, George would have answered correctly," which is not obviously true. (Weiss 2017, p.390)

Let us denote the relevant counterpossibles.

(1) If 5+7 were 13, then George would have answered correctly.

(2) If 5+7 = 13, then 5+6 = 12.

(3) If 5+6 = 12, then 5+6 = 13.

(4) If 5+6 were 12, then George would have answered correctly.

Both (2) and (3) seem true enough, although not as obviously as (1) does. Weiss discounts all potential objections attacking the soundness of the argument as question begging on the basis of the intuitive truth of (2) and (3). This riposte has some merit but seems a little too quick, and as such introduces problems of its own. One way of arguing against their truth is to say that contexts where we want (1) to be true, need not always be ones where we'd be also willing to admit (2) and (3) as true. Indeed, there seem to be many ways of arguing against the truth of (2) and (3) in contexts where (1) is true, however I agree that it doesn't seem obvious that there should be no context at all where we would allow all three to be true, thereby admitting the counterexample as legitimate.

However, caution should be exercised when accepting a general strategy for generating counterexamples that admits the truth of premises whose relevance to the pertinent context can be questioned, because this may pave the way to invalidating more than one has bargained for. Finally, Weiss discounts all potential counter-objections that would defend the truth of (4), by asserting that it is intuitively false. But to me it doesn't seem all that much less acceptable than what is already taken on-board when admitting both (2) and (3) as true. As a matter of fact, by admitting those additional premises, (4) doesn't seem as odd as it would be in their absence.[195]

The first thing to note is that (1) bares very close resemblance to a statement of counterpossible identity, and as such is intuitively true. That is, it appears to mean no more and no less than:

(1.a) If 5+7 were 13, then George answering '13' to the question what '5+7' is, would have answered correctly.

[195] Weiss uses a Sorites kind of reasoning, which employs numerous applications of (5.10), to amplify the salience of the falsehood of (4) further (or rather, diminish any intuitive claim to truth that (4) may have) and derive an arguably much less plausible version (4'). But a similar objection can be set against it, i.e. that the amplified conclusion (4') is no less obvious than the assumed stability of reasoning involved in deriving it and the truth of all the intermediate steps required to arrive at (4').

Or even more explicitly:

(1.b) 'If 5+7 were 13, then George saying '5+7 is 13' would be telling the truth'

We're dealing with a counterpossible whose consequent's impossible content is not ampliative relative to the antecedent, i.e. no greater than the content of (1)'s antecedent. The consequent can be said to very naturally follow from the antecedent. Or, yet to put it another way, which (1.b) intends to highlight, (1) is closely related to a relatively safe counter-factual/possible:

(1.c) If *A* were true, then saying '*A*' would be to speak truly.

Therefore, worlds where the consequent is true would certainly seem like scenarios no stranger than those corresponding to whatever (impossibility) is expressed in the antecedent alone. However, this doesn't seem to be the case for the admission of the truth of (2), which, on top of the impossibility expressed in the antecedent requires us to accept additional assumptions about arithmetic in such impossible situations, which feels *stranger* than the scenario envisaged in (1). That would speak in support of an argument that the context has indeed shifted – but we're allowing for that, so let's continue. In other words, granting the truth of (2), and then (3) would seem to be stretching the strangeness of a world *w* that suffices to make (1) true. So, the antecedent world or worlds required to make (2) and (3) true, should at least be distinct from the antecedent world *w* that makes (1) true and certainly no more similar to the actual world than *w*. Those would seem to be the correct and weakest comparative similarity requirements fitting this scenario. With these assumptions about comparative similarity in place, the task is to salvage the truth of (1), (2), and (3) without committing to the truth of (4). This is impossible on a notion of comparative similarity of worlds with unrestricted comparability of worlds, i.e. characterized by total preorders. But perhaps it could be argued that the antecedent world (or worlds) required for the non-vacuous truth of (2) and (3) is not so much *stranger* than the antecedent world(s) required for the non-vacuous truth of (1) but *strange in a different way*. This interpretation of comparative similarity/dissimilarity of worlds could potentially serve as intuitive motivation for abandoning comparability over impossible worlds.

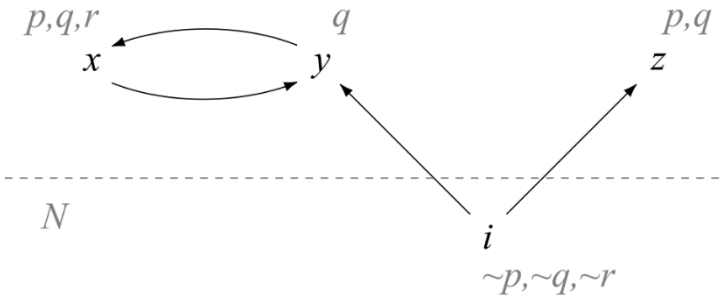
5.4.2 Weaker totality condition (T1)

Totality of $\precsim_i$ need not be abandoned *altogether* to avoid commitment to (5.10). It suffices to lift totality from impossible worlds only, i.e. by replacing the stronger comparability

condition (T2) with a weaker one (T1). Note that systems that invalidate (5.10), by virtue of satisfying the weaker comparability condition (T1) aren't somehow particularly contrived. At least, they are not any more contrived than some aspects of the model theory that are already in place. That is, we've already distinguished the elements of N and $W\backslash N$ at the level of models, by stipulating distinct conditions for ρ (and consequently $\Vdash$). So, it doesn't seem all that more contrived to distinguish the elements of N and $W\backslash N$ at level of ordering frames, by allowing distinct conditions for $\lesssim$.

Proposition 5.2.1: $p > q, q > p \nVdash_{\mathbf{CS}_3^*} (p > r) \equiv (q > r)$

Proof : Let $\mathfrak{A} = (W, N, \{\lesssim_i : i \in N\}, \rho)$, be a $\mathbf{CS}_3^*$ model such that $W = \{i, x, y, z\}$, $N = \{i\}$, the following ordering assignment $\lesssim_i = \{(i,x), (i,y), (i,z), (x,y), (y,x), (i,i), (x,x), (y,y), (z,z)\}$, and $\Vdash = \{(i, \sim p), (i, \sim q), (i, \sim r), (x,p), (x,q), (x,r), (y,q), (z,p), (z,q)\}$. See diagram below (indication of reflexivity has been omitted for better readability).



Now, $z \in [p] \cap [q] = \{x, z\}$, and $\downarrow z \cap [p] = \{z\} \subseteq [q] = \{x, y, z\}$, and $\downarrow z \cap [q] = \{z\} \subseteq [p] = \{x, z\}$. Hence $i \Vdash \{p > q, q > p\}$. Also $x \in [p] = \{x, z\}$ and $\downarrow x \cap [p] = \{x\} \subseteq [r] = \{x\}$, therefore $i \Vdash p > r$. But, $\downarrow w \cap [q] \not\subseteq [r]$, for all $w \in [q]$, i.e. $\downarrow x \cap [q] = \downarrow y \cap [q] = \{x, y\} \not\subseteq [r] = \{x\}$, and $\downarrow z \cap [q] = \{z\} \not\subseteq [r] = \{x\}$. Hence $i \nVdash q > r$, as required. □

5.4.3 (T1) and *Adjunction of Consequents*

Now we turn to another motivation for (T1). Lifting unrestricted comparability over impossible worlds is a way of invalidating inferences that have at least one pair of counterpossible premises with the same antecedent, and whose validity hinges on all counterpossibles with the same antecedents being evaluated on the same relevant set of worlds. So, for example, inferences like (5.11) – namely, *Adjunction of Consequents*, discussed extensively in §4 – will be valid in all conditional logics where the antecedent is the only parameter that determines the range of the accessibility relation.

(5.11) $A > B, A > C \vDash A > (B \wedge C)$

This is formally valid on any labelled transition system model $(W, N, \{R_A : A \in For\}, \rho)$ where ρ and $\Vdash$ are defined exactly the same as in definition 5.3, where $(W, \{R_A : A \in For\})$, R_A, and $f_A(w)$ are as in §2.1.3, and the truth conditions for $>$ are: $w \Vdash A > B$ iff $f_A(w) \subseteq [B]$.[196]

But consider the following instance (which appears to be a clear counterexample) containing Goodman-inspired counteridenticals:[197]

(1) If the number 2 was Sherlock Holmes, then 2 would be a detective.
(2) If Sherlock Holmes was the number 2, then Sherlock Holmes wouldn't be a detective.
(3) Therefore, if 2 was S. Holmes, then 2 would be a detective and S. Holmes wouldn't be a detective.

Both premises are non-vacuously true counterpossibles, however the conclusion (3) is clearly not true.[198] CS* systems with (T1) instead of (T2) give a correct analysis of this counterexample and ones like it, i.e. there are countermodels to (5.11) in each system weaker than $\mathbf{CS}_4^*$, and I explicitly give one in *Proposition 5.3*, below. One may object to admitting both premises, on account of apparent radical context-shift required for that, but then one would have to decide how much of a context shift between premises is allowed. At least it's not obviously clear that the freedom of context-shift employed in the counterexample to (5.10), discussed earlier, would not justify the context shifts in the invalidation of (5.11). And (5.11) is formally valid in all the alternative non-vacuist systems that Weiss (2017) endorses.[199]

The inference is still valid for systems with the stronger comparability condition (T2).

Proposition 5.3: $A > B, A > C \vDash_{\mathbf{CS}_4^*} A > (B \wedge C)$

Proof: Let $\mathfrak{A} = (W, N, \{\lesssim_i\}_{i \in N}, \rho)$, be a $\mathbf{CS}_4^*$ model and let $i \Vdash \{A > B, A > C\}$ for arbitrary $i \in W$. Then $\exists k \in [A]$ such that $\downarrow k \cap [A] \subseteq [B]$, and $\exists k' \in [A]$ such that $\downarrow k' \cap [A] \subseteq [C]$. Now,

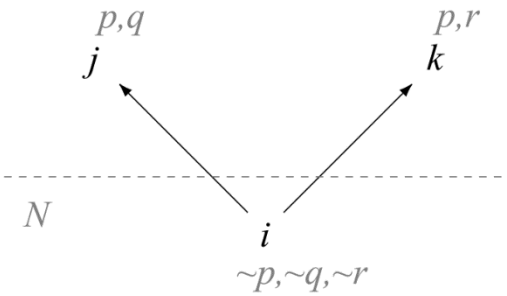
[196] This follows from *Proposition 4.12*, (and footnote 168) which shows that (5.11) is valid for **CS** and C.
[197] Goodman (1983, p.6). Note that (1) and (2) are equivalent – the order has been inverted only for emphasis.
[198] I'm assuming that being a detective is an essential property of Sherlock Holmes (i.e. Holmes has it in all possible worlds), and the number 2 is not a detective in any possible world, hence the identity is a counterpossible identity.
[199] Although Weiss builds the alternative non-vacuist proposal on conditional logics like C and C+ (i.e. labelled transition systems that take only the antecedent as a parameter in the accessibility relation), he does entertain similar systems, which also include the consequent as a parameter in characterizing the accessibility relation, drawing on Gabbay's (1972) proposal – an approach that would allow for the invalidation of (5.11).

either $k \lesssim_i k'$ or $k' \lesssim_i k$ by (T2). If $k \lesssim_i k'$, then clearly $\downarrow k \subseteq \downarrow k'$, which in conjunction with the hypothesis implies $\downarrow k \cap [A] \subseteq [C]$. So, we have both $\downarrow k \cap [A] \subseteq [B]$ and $\downarrow k \cap [A] \subseteq [C]$, which jointly imply $\downarrow k \cap [A] \subseteq [B] \cap [C]$. Hence $i \Vdash A > (B \wedge C)$, as required. A very similar argument holds for the case when $k' \lesssim_i k$. □

However, it fails once we weaken the comparability condition to (T1).

Proposition 5.3.1: $p > q, p > r \nvDash_{\mathbf{CS}_3^*} p > (q \wedge r)$

Proof : Let $\mathfrak{A} = (W, N, \{\lesssim_i : i \in N\}, \rho)$, be a $\mathbf{CS}_3^*$ model such that $W = \{i, j, k\}$, $N = \{i\}$, letting $\lesssim_i = \{(i,j), (i,k), (i,i), (j,j), (k,k), \}$, and $\Vdash = \{(i, \sim p), (i, \sim q), (i, \sim r), (j, p), (j, q), (k, p), (k, r)\}$. See diagram below (indication of reflexivity has been omitted for better readability).



Now, $j \in [p]$ and $\downarrow j \cap [p] = \{j\} \subseteq [q] = \{j\}$, and also $k \in [p]$ and $\downarrow k \cap [p] = \{k\} \subseteq [r] = \{k\}$. Hence, $i \Vdash \{p > q, p > r\}$. But $\downarrow w \cap [p] \nsubseteq [q \wedge r] = \emptyset$, for all $w \in [p]$. So, $i \nVdash p > (q \wedge r)$, as required. □

5.4.4 (T1) and mere counterfactuals

The benefits of (T1) are not confined to counterpossibles. That is, all CS* systems weaker than $\mathbf{CS}_4^*$ invalidate (5.11) even when it is confined to counterfactuals. As exemplified in §4.5 there exist intuitive counterexamples to (5.11) confined to mere counterfactuals, which nevertheless go through in C and CS because the premises can never be jointly true at any possible world (see claim 4.5.2 and f.21 in §4.5). That is, and this is a crucial point (reiterated from §4.5), the counterexamples go through vacuously because the *combined truth of both counterfactual premises* at some world – particularly in cases of said counterexamples – *implies inconsistent situations*, once the imported information is accounted for, and so if the analysis is restricted to possible worlds, the premises can't be jointly true (since inconsistent situations can't be accommodated). But on CS* systems with the weaker comparability condition (T1), there is a way of preserving all the intuitive scope of (5.11)'s applicability (it

only fails in cases when the truth of the premises depends on radical context shifts), and at the same time allow it to fail by giving an accurate analysis of the counterexamples. This is achieved by:

(i) Accommodating the implied inconsistent situations – necessitated by the truth of both premises – by allowing them to hold at impossible worlds.

(ii) Lifting comparability from impossible worlds to block the truth of the conclusion.

That is, the inference is analysed as invalid for counterfactuals, as it should be, for the same general reasons that motivate non-vacuism. Therefore, returning to the counterexample discussed at length in §4.5, there is a CS* model where (1) and (2) are true at the actual world, and (3) is false, as required – just take *Proposition 5.3.1* as the corresponding countermodel.

(1) If Everest was in New Zealand, Everest would be in the Southern Hemisphere.

(2) If Everest was in New Zealand, New Zealand would be in the Northern Hemisphere.

(3) Therefore, if Everest was in New Zealand, then Everest would be in the Southern Hemisphere and New Zealand would be in the Northern Hemisphere.

Lifting totality does seem like a substantial change, even if restricted to impossible worlds, and some more reflection is required. Nevertheless (5.12), which is a (SI2) salvaged version of (5.10) is valid in $\mathbf{CS}_3^*$. A similar solution will not work for (5.11), since the antecedent in cases where the inference fails need not express an impossibility, e.g. Mount Everest being in New Zealand is a perfectly possible scenario.

(5.12) $\Diamond A, \Diamond B, A > B, B > A \vDash (A > C) \equiv (B > C)$

Proposition 5.4: $\Diamond A, \Diamond B, A > B, B > A \vDash_{\mathbf{CS}_3^*} (A > C) \equiv (B > C)$

Proof : It suffices to observe that the premises, if true, in conjunction with (SI2) will not require the evaluation of $>$ formulae to access any impossible worlds. Only possible worlds will be accessed – and those are totally preordered by (T1). So, the inference goes through, since we need failure of comparability for a counterexample, by *proposition 5.2.1*, but with (T1) in place only impossible worlds can be incomparable. □

Conclusion

In the culminating chapters (4 and 5) of this thesis I have shown that there are accessible modifications of Lewis' (1981) ordering semantics for analyses of counterfactuals that resolve some persistent issues regarding contextual ambiguity, and avoid the inadequacy of treating all counterpossibles as vacuously true or false. Moreover, in each case it has been indicated to what extent each modification preserves the logic that serves as its basis.

The account of context-indexed counterfactuals, proposed in chapter 4, not only addresses the context related concerns identified by Goodman (1954) and Quine (1966), but also offers a meaningful notion of semantic consequence based on the idea of contextual information preservation. Although the approach chosen may not be optimal, given that it also modifies the object language, nevertheless there is some evidence that it appears to be a natural move. But even if the offered account is vulnerable to the charge of not providing a direct solution to the pertinent contextual issues – as it departs from the analysis of a single conditional – at least it offers a framework of thinking about those issues, which can be viewed as going some way toward providing such a solution.

The account of counterfactuals, proposed in chapter 5, avoids a vacuist account of counterpossibles, whilst preserving much of Lewis' analysis of mere counterfactuals. The application of an impossible world semantics in this case has been partly motivated and justified by the redeeming roles that such semantics have played in other areas of philosophical analysis and logic. However, although I have replied to Lewis' defense of vacuism and his objection to impossible worlds, I admit that I have not given a comprehensive defense to all the existing objections. I have defended the feasibility of the comparative similarity of worlds interpretation of impossible world ordering semantics for counterfactuals against some recent objections, by showing that counterexamples arming such objections can be invalidated on systems where impossible worlds satisfy a weaker comparability condition, i.e. partial preorderhood. However, there do remain other questions regarding some key ordering conditions, underlying the extended domain that could be addressed in the future.

A natural step would be to combine the results from chapters 4 and 5, and fashion a system that gives both an adequate analysis of counterpossibles, whilst accounting for contextual distinctions. That is, we can employ the benefits of the system **CS2**+ and modify its definition so it is based on **CS*** models instead of (vacuism-burdened) **CS** models, noting that $\mathbf{CS}_3^*$ has been argued to be the optimal **CS*** system. This way we would have a system that inherits the advantages of both **CS2**+ and $\mathbf{CS}_3^*$. This and a more comprehensive participation in the defense of impossible world semantics in general would constitute a well-motivated inclusion to future research.

Appendix

The following appendix contains the proof of the equivalence of the class of **CS** models and the class of **S** models. That is, these classes validate the same sets of formulae and inferences. This proof is based on a proof sketch given by Lewis (1973, p.49). First, let us recall the relevant definitions, so the formulation of the theorem is clear.

Definition 2.15: Let $\vDash_{\mathbf{S}} \subseteq \wp(For) \times For$, and define $\Sigma \vDash_{\mathbf{S}} A$ iff for all models $(W, \$, [\ ])$, and all $w \in W$, if $w \Vdash B$ for all $B \in \Sigma$, then $w \Vdash A$. That is, valid inference is defined as truth preservation at all worlds in all systems of spheres models. A formula $A \in For$ is said to be valid iff $\emptyset \vDash_{\mathbf{S}} A$. Call this logic **S**.[200]

Definition 4.2.7: Let $\vDash_{\mathbf{CS}} \subseteq \wp(For) \times For$, and define $\Sigma \vDash_{\mathbf{CS}} A$ iff for all models $(W, \lesssim, \rho)$, and all $i \in W$, if $i \Vdash B$ for all $B \in \Sigma$, then $i \Vdash A$. We say an inference from Σ to A is valid iff $\Sigma \vDash_{\mathbf{CS}} A$. That is, valid inference is defined as truth preservation at all worlds in all **CS**-models. A formula $A \in For$ is said to be valid iff $\emptyset \vDash_{\mathbf{CS}} A$. Call this logic **CS**.

Theorem *A*.1.1: $\Sigma \vDash_{\mathbf{S}} A$ iff $\Sigma \vDash_{\mathbf{CS}} A$

Proof : First construct injective maps $f: \mathbf{CS} \longrightarrow \mathbf{S}$ and $g: \mathbf{S} \longrightarrow \mathbf{CS}$ between the class of CS frames and S frames (definitions *A*.4.1, *A*.4.2), such that (i) for each CS frame $\mathfrak{F}$, $f(\mathfrak{F})$ is an S frame (lemma *A*.1.0.1) and (ii) for any S frame $\mathfrak{F}$, $g(\mathfrak{F})$ is a CS frame (lemma *A*.1.0.2).[201]

Then, show both f and g to be truth preserving in the following sense (lemmas *A*.1.0.4, *A*.1.0.4):
For any **CS** frame $\mathfrak{F} = (W, \lesssim)$, $i \in W$, ρ, and $A, B \in For$:

$(\mathfrak{F}, \rho), i \Vdash A > B$ iff $(f(\mathfrak{F}), \rho), i \Vdash A > B$

For any **S** frame $\mathfrak{F} = (W, \$)$, $i \in W$, ρ, and $A, B \in For$:

$(\mathfrak{F}, \rho), i \Vdash A > B$ iff $(g(\mathfrak{F}), \rho), i \Vdash A > B$

Since the above also hold for the basic modal language, the result follows.

---

[200] For clarity of presentation I should redefine S models in terms of S frames and ρ rather than [ ]. Keeping track of that irrelevant, yet nontrivial difference would be an unnecessary distraction.
[201] I follow Lewis' (1973, p.49) proof idea here – in particular, the definitions of the injections *f* and *g*.

For the purposes of the following, let's recall the (relevant) precise definition of the general notion of arbitrary unions and arbitrary intersections. Given a collection of sets $\boldsymbol{\mathcal{S}}$:

$$x \in \bigcup \boldsymbol{\mathcal{S}} \iff \exists A \in \boldsymbol{\mathcal{S}}, x \in A. \qquad x \in \bigcap \boldsymbol{\mathcal{S}} \iff \forall A \in \boldsymbol{\mathcal{S}}, x \in A.$$

Definition *A*.1.1: Define the following map: $f\colon \mathbf{CS} \longrightarrow \mathbf{S}$ as follows:

$f(W) = W.$

$f\big((S_i, \lesssim_i)\big) := \$_i^{\lesssim} = \big\{S \in \wp(S_i)\colon (\forall j, k \in W)\big((j \in S \wedge k \notin S) \longrightarrow j <_i k\big)\big\}$

It may be helpful to think of elements of $\$_i^{\lesssim}$ as downward $\lesssim_i$-closed sets. That is, $S \in \$_i^{\lesssim}$ iff $(j \in S \wedge k \lesssim_i j) \longrightarrow k \in S$ for any $\forall j, k \in W$.

Definition *A*.1.2: Define the following map: $g\colon \mathbf{S} \longrightarrow \mathbf{CS}$ as follows:

$g(W) = W.$

$g(\$_i) := \big(\lesssim_i^{\$}, S_i^{\$}\big)$

- $\lesssim_i^{\$} = \{(j, k) \in W \times W\colon (\forall S \in \$_i)(k \in S \longrightarrow j \in S)\}$
- $S_i^{\$} = \bigcup \$_i$

Lemma *A*.1.0.1: For each **CS** frame $\mathfrak{F} = (W, \lesssim)$, $f(\mathfrak{F}) = (W, \$^{\lesssim})$ is an **S** frame.

Proof : Let $(W, \lesssim)$ be **CS** frame (as per definition 4.22). We want to show that each $(W, \$^{\lesssim})$ is an **S** frame, i.e. that $\$_i^{\lesssim}$ satisfies *nesting* and *weak centering*, for each $i \in W$.

Nesting: for all $S, T \in \$_i^{\lesssim}$ either $S \subseteq T$ or $S \subseteq T$. This follows from totality of $\lesssim$. For the trivial case, suppose $S, T \in \$_i^{\lesssim}$, and let $S = \emptyset$. Hence, $S \subseteq T$. Now, suppose for contradiction that there are nonempty sets $S, T \in \$_i^{\lesssim}$ such that neither $S \subseteq T$ nor $T \subseteq S$. First, suppose that it's not the case that $S \subseteq T$. So, there is some $x \in S_i$ such that $x \in S$ but $x \notin T$. Next, also assume that it's not the case that $T \subseteq S$. So, there is some $z \in S_i$ such that $z \in T$ but $z \notin S$. Hence, from $x \in S$, $z \notin S$, and the definition of $\$_i^{\lesssim}$ we infer $x <_i z$, and from $z \in T, x \notin T$, and the definition of $\$_i^{\lesssim}$ and we also infer $z <_i x$. But this is impossible, since $x <_i z$ and $z <_i x$ means $(z, x) \notin \lesssim$ and $(x, z) \notin \lesssim$, by definition of $<_i$, which contradicts totality of $\lesssim$.

Weak Centering: $\exists S \in \$_i^{\lesssim}(S \neq \emptyset)$ and $i \in \bigcap\big(\$_i^{\lesssim} \setminus \emptyset\big)$. First to show $\exists S \in \$_i^{\lesssim}(S \neq \emptyset)$. It suffices to observe that, by definition, $\lesssim_i$ satisfies CS5: $\forall j, k \in W\big((j \in S_i \wedge k \notin S_i) \longrightarrow j <_i k\big)$. So, $S_i \in \$_i^{\lesssim}$. Also $S_i \neq \emptyset$, since $i \in S_i$, by CS2. Next, suppose for contradiction that there is some

nonempty sphere $S \in \$_i^{\lesssim}$ such that $i \notin S$. Now $S \neq \emptyset$ implies that there is some $j \in W$ such that $j \in S$. Hence, in particular $(j \in S \wedge i \notin S) \longrightarrow j <_i i$ is true, by $\$_i^{\lesssim}$ membership. But since the antecedent is true by hypothesis, it follows that $j <_i i$ for some $j \in W$. But this is impossible, since it contradicts CS3, i.e. the $\lesssim_i$-minimality of i. This completes the proof. □

Lemma *A*.1.0.2: For each **S** frame $\mathfrak{F} = (W, \$)$, $g(\mathfrak{F}) = \left(W, \lesssim^{\$}\right)$ is a **CS** frame.
Proof : Let $(W, \$)$ be an **S** frame (as per definition 2.17). We want to show that $\left(W, \lesssim^{\$}\right)$ is a **CS** frame. First, to show that each $\left(S_i^{\$}, \lesssim_i^{\$}\right)$ is a *total preorder* for each $i \in W$, i.e. it satisfies CS1.

Transitivity: $\forall x, y, z \in W \left((x \lesssim_i^{\$} y \wedge y \lesssim_i^{\$} z) \longrightarrow x \lesssim_i^{\$} z\right)$. Suppose $(x, y) \in \lesssim_i^{\$}$ and $(y, z) \in \lesssim_i^{\$}$ for any $x, y, z \in S_i^{\$}$. From the definition of $\lesssim_i^{\$}$, this implies $y \in S \longrightarrow x \in S$ and $z \in S \longrightarrow y \in S$ for all $S \in \$_i$. Hence, $z \in S \longrightarrow x \in S$ for all $S \in \$_i$, by hypothetical syllogism, and $(x, z) \in \lesssim_i^{\$}$ by definition of $\lesssim_i^{\$}$.

Totality: $\forall x, y \in W \left(x \lesssim_i^{\$} y \vee y \lesssim_i^{\$} x\right)$. Suppose for contradiction that there are $x, y \in S_i^{\$}$ such that $(x, y) \notin \lesssim_i^{\$}$ and $(x, y) \notin \lesssim_i^{\$}$. From the definition of $\lesssim_i^{\$}$, this implies that there exist spheres $S, T \in \$_i$ such that $y \in S \wedge x \notin S$ and $x \in T \wedge y \notin T$. Now, $S \subseteq T$ is impossible, because $y \in S$ but $y \notin T$. By nesting, the only other possibility is $T \subsetneq S$. Suppose $T \subsetneq S$, but that's also impossible because $x \in T$ but $x \notin S$.

Next, to show that each $\lesssim_i^{\$}$ satisfies the remaining conditions CS2-CS5.
(CS2) The world i is self-accessible: $i \in S_i$. To show $i \in S_i^{\$}$. Since $\$_i$ is weakly centered, there is a sphere $\emptyset \neq S \in \$_i$, such that $S = \bigcap \$_i$ and $i \in \bigcap \$_i \subseteq \bigcup \$_i = S_i^{\$}$, as required.
(CS3) The element i is $\lesssim_i$-minimal: $\forall j \in W (i \lesssim_i j)$. Suppose for contradiction that there exists $i \neq j \in S_i^{\$}$ such that $(i, j) \notin \lesssim_i^{\$}$. So, by definition of $\lesssim_i^{\$}$ there's a $T \in \$_i$ such that $j \in T \wedge i \notin T$. This contradicts weak centering, which requires that i is included in every nonempty sphere in $\$_i$.

(CS4) Inaccessible worlds are $\lesssim_i^{\$}$-maximal: $\forall j, k \in W\left(k \notin S_i^{\$} \longrightarrow j \lesssim_i^{\$} k\right)$. It suffices to observe that each inaccessible world $k \notin \bigcup \$_i = S_i^{\$}$ satisfies the above condition, by definition of $\lesssim_i^{\$}$. That is, $(j, k) \in \lesssim_i^{\$}$ for each $k \notin S_i^{\$}$ and any $j \in W$, since $(\forall S \in \$_i)(k \in S \longrightarrow j \in S)$ is satisfied vacuously for all inaccessible worlds (worlds outside of $\bigcup \$_i$), i.e. the antecedent is always false.

(CS5) Accessible worlds are more similar to i than inaccessible worlds:

$$\forall j,k \in W \left(\left(j \in S_i^{\$} \wedge k \notin S_i^{\$} \right) \rightarrow j <_i^{\$} k \right)$$

Suppose for contradiction that $k \notin S_i^{\$}$ but $(k,j) \in \lesssim_i^{\$}$ for some $j \in S_i^{\$}$. This means that $k \notin S$ for all $S \in \$_i$ and that there is some $T \in \$_i$ such that $j \in T$. Now, $(k,j) \in \lesssim_i^{\$}$ implies $j \in S \rightarrow k \in S$ for all $S \in \$_i$, by definition of $\lesssim_i^{\$}$. In particular $j \in T \rightarrow k \in T$, implying $k \in T$, which is impossible.

This completes the proof. □

Lemma *A*.1.0.3: Functions f and g, as given in definitions A.1.1 and A.1.2, are *injections*.

Proof : It is immediate from definitions. □

There's a pattern to all the proof directions in lemmas *A*.1.0.4 and *A*.1.0.5.

Regarding the non-trivial cases, we're dealing with two kinds of conditions (ii) and (II), that vary slightly, but have the same general form. Hence the proofs follow a similar pattern. Below I give formulations that aims to emphasize the similarity of the forms of the non-vacuous conditions.

Ordering frames (**): $(\exists x)\left(P(x) \wedge (\forall y)\left(R(y,x) \rightarrow Q(y)\right)\right)$

Similarity spheres (##): $(\exists X)\left(P'(X) \wedge (\forall y)\left(R'(y,X) \rightarrow Q(y)\right)\right)$

All the proofs (with some variation) follow a general pattern:

(**) ⟶ (##):

(1) First, I show that $(\exists x)P(x)$ gives $(\exists X)P'(X)$.

(2) Next, I show that $(\forall y)\left(R'(y,X) \rightarrow R(y,x)\right)$.

Steps (1) and (2) generally require the most work, and I use various methods.

(3) I use (2) in conjunction with the second conjunct of the hypothesis $(\forall y)\left(R(y,x) \rightarrow Q(y)\right)$ to show that $(\forall y)Q(y)$, by hypothetical syllogism, thus proving $(\forall y)\left(R'(y,X) \rightarrow Q(y)\right)$, by conditional proof.

For the (##) ⟶ (**) direction, I use the same proof pattern as for (**) ⟶ (##).

Lemma *A*.1.0.4: For any **CS** frame $\mathfrak{F} = (W, \lesssim)$, $i \in W$, ρ, and $A, B \in For$:

$$(\mathfrak{F}, \rho), i \Vdash A > B \quad \text{iff} \quad (f(\mathfrak{F}), \rho), i \Vdash A > B$$

Proof: Suppose $\big((W,\lesssim),\rho\big), i \Vdash A > B$ for arbitrary $(W,\lesssim) \in \mathbf{CS}$, $i \in W$, ρ, and $A, B \in For$ Then by definition, either

(i) $S_i \cap [A] = \emptyset$, or

(ii) $(\exists k)\big(k \in S_i \cap [A] \ \wedge\ (\forall j \in W)(j \lesssim_i k \ \longrightarrow j \in [A \supset B])\big)$

We need to show that either

(I) $\cup \$_i^{\lesssim} \cap [A] = \emptyset$, or

(II) $\big(\exists S \in \$_i^{\lesssim}\big)\big(S \cap [A] \neq \emptyset \ \wedge\ (\forall j \in W)(j \in S \longrightarrow j \in [A \supset B])\big)$

That is, we need to show that (i or ii) iff (I or II). The entire argument applies for any $i \in W$.

(i $\leftrightarrow$ I) The vacuous case: $\cup \$_i^{\lesssim} \cap [A] = \emptyset$ iff $S_i \cap [A] = \emptyset$, since $\cup \$_i^{\lesssim} \subseteq S_i$, by definition of $\$_i^{\lesssim}$, and $S_i \in \$_i^{\lesssim}$, by definition of $\lesssim_i$ and $\$_i^{\lesssim}$, i.e. CS5 implies $S_i \in \$_i^{\lesssim}$.

(ii $\longrightarrow$ II) Assume there is a $k \in S_i \cap [A]$. I'll now define a subset of $K \subseteq W$, such that $k \in K$, and show that $K \in \$_i^{\lesssim}$. In other words, I'll define a set $K \subseteq W$ whose existence is implied by the existence of k. Let $K := \{j \in W : j \lesssim_i k\}$. Observe that $K \subseteq S_i$, since $k \in S_i$ and $j \lesssim_i k$ jointly imply $j \in S_i$ for all $j \in W$ (S_i is downward $\lesssim_i$-closed). Denying this would contradict CS5, i.e. suppose there's some $j \in W$ such that $j \lesssim_i k$ yet $j \notin S_i$. But $k \in S_i$ and $j \notin S_i$ implies $k <_i j$, by CS5, contradicting $j \lesssim_i k$. Now I'll show that $K \in \$_i^{\lesssim}$. It suffices to note that $K \in \$_i^{\lesssim}$ follows from K being downward $\lesssim_i$-closed, i.e. for any $x, y \in W$, if $x \in K$ and $y \lesssim_i x$, then $y \in K$, which can be easily checked as being equivalent to $(x \in K \wedge y \notin K) \longrightarrow x <_i y$ for any $x, y \in W$. Hence, $K \in \$_i^{\lesssim}$, by definition of $\$_i^{\lesssim}$. Now I'll show that $j \in K \longrightarrow j \in [A \supset B]$ for all $j \in W$. Suppose $j \in K$ for arbitrary $j \in W$. Hence, $j \lesssim_i k$, by construction of K. Next, from hypothesis $j \lesssim_i k \ \longrightarrow j \in [A \supset B]$ for all $j \in W$, we conclude $j \in [A \supset B]$. Hence, $j \in K \longrightarrow j \in [A \supset B]$ for all $j \in W$, by conditional proof, as required.

(II $\longrightarrow$ ii) Assume that there is a sphere $S \in \$_i^{\lesssim}$ such that $S \cap [A] \neq \emptyset$, i.e. there's some $k \in S \cap [A]$. So, there is a $k \in S_i \cap [A]$, since $\cup \$_i^{\lesssim} \subseteq S_i$, by definition of $\$_i^{\lesssim}$. Now I'll show that $j \lesssim_i k \longrightarrow j \in S$ for all $j \in W$. To that end it suffices to note that $x \lesssim_i y \longrightarrow (y \in T \longrightarrow x \in T)$ for any $x, y \in W$ and $T \in \$_i^{\lesssim}$ is the contraposed condition for $\$_i^{\lesssim}$ membership. Now, assume $j \lesssim_i k$ for arbitrary $j \in W$. Hence, $k \in T \longrightarrow j \in T$ for any $T \in \$_i^{\lesssim}$ and $j \in W$. In particular $k \in S \longrightarrow j \in S$ for any $j \in W$, by hypothesis $S \in \$_i^{\lesssim}$. Therefore $j \in S$ for all $j \in W$ that satisfy $j \lesssim_i k$, by hypothesis $k \in S$. Hence, $j \lesssim_i k \longrightarrow j \in S$ for all $j \in W$, by conditional proof. This, in conjunction with the main hypothesis $(\forall j \in W)(j \in S \longrightarrow j \in [A \supset B])$ in (II) gives $j \in [A \supset B]$

for all $j \in W$, by hypothetical syllogism. Hence, finally $j \lesssim_i k \longrightarrow j \in [A \supset B]$ for all $j \in W$, by conditional proof, as required. □

Lemma *A*.1.0.5: For any **S** frame $\mathfrak{G} = (W, \$)$, $i \in W$, ρ, and $A, B \in For$:

$$(\mathfrak{F}, \rho), i \Vdash A > B \quad \text{iff} \quad (g(\mathfrak{F}), \rho), i \Vdash A > B$$

Proof: Suppose $\big((W, \$), \rho\big), i \Vdash A > B$ for arbitrary $(W, \$) \in \mathbf{S}$, $i \in W$, ρ, and $A, B \in For$
Then by definition, either

(i) $\bigcup \$_i \cap [A] = \emptyset$, or

(ii) $(\exists S \in \$_i)\big(S \cap [A] \neq \emptyset \ \wedge \ (\forall j \in W)(j \in S \longrightarrow j \in [A \supset B])\big)$

We need to show that either

(I) $S_i^{\$} \cap [A] = \emptyset$, or

(II) $(\exists k)\Big(k \in S_i^{\$} \cap [A] \ \wedge \ (\forall j \in W)\big(j \lesssim_i^{\$} k \ \longrightarrow j \in [A \supset B]\big)\Big)$

That is, we need to show that (i or ii) iff (I or II). The entire argument applies for any $i \in W$.

(i $\leftrightarrow$ II) The vacuous case is immediate, since $S_i^{\$} = \bigcup \$_i$ by definition of g.

(ii $\longrightarrow$ II) Assume that there is a $S \in \$_i$ such that $S \cap [A] \neq \emptyset$, i.e. there is a $k \in S \cap [A]$. So, there's a $k \in S_i^{\$} \cap [A]$, since $S \subseteq S_i^{\$} = \bigcup \$_i$ by definition of g. Now I'll show that $j \lesssim_i^{\$} k \longrightarrow j \in S$ for any $j \in W$. Suppose $j \lesssim_i^{\$} k$ for arbitrary $j \in W$. Hence, $k \in T \longrightarrow j \in T$ for any and $T \in \$_i$ and $j \in W$, by definition of $\lesssim_i^{\$}$, so in particular $k \in S \longrightarrow j \in S$ for any $j \in W$, by hypothesis $S \in \$_i$. So, $j \in S$ for all $j \in W$ that satisfy $j \lesssim_i^{\$} k$, by hypothesis $k \in S$. Hence, $j \lesssim_i^{\$} k \longrightarrow j \in S$ for all $j \in W$, by conditional proof. This, in conjunction with main hypothesis $(\forall j \in W)(j \in S \longrightarrow j \in [A \supset B])$ in (ii) gives $j \in [A \supset B]$ for all $j \in W$, by hypothetical syllogism. Hence $j \lesssim_i^{\$} k \longrightarrow j \in [A \supset B]$ for all $j \in W$, by conditional proof, as required.

(ii $\longleftarrow$ II) Assume there is a $k \in S_i^{\$} \cap [A]$, so $\bigcup \$_i \cap [A] \neq \emptyset$, by definition of g, and there is a sphere $S \in \$_i$, such that $k \in S \cap [A]$, by definition of set union. Now we need to show that if $j \in S$, then $j \lesssim_i^{\$} k$ for any $j \in W$. By definition of $j \lesssim_i^{\$} k$, $k \in T \longrightarrow j \in T$ for all $T \in \$_i$, so in particular $k \in S \longrightarrow j \in S$. Hence $j \in S$. Hence $j \in S \longrightarrow j \lesssim_i^{\$} k$, for any $j \in W$, by conditional proof. Finally, in conjunction with $(\forall j \in W)\big(j \lesssim_i^{\$} k \ \longrightarrow j \in [A \supset B]\big)$ from main hypothesis (II), we conclude that $j \in [A \supset B]$ for any $j \in W$. Hence $(\forall j \in W)(j \in S \ \longrightarrow j \in [A \supset B])$, by conditional proof. □

References

Berto, F. (2010). Impossible worlds and propositions: Against the parity thesis. *The Philosophical Quarterly*, *40*, 471–86.

Berto, F. (2017). Impossible worlds and the logic of imagination. *Erkenntnis*, *82*, 1277-1297.

Berto, F. (2014). On conceiving the inconsistent. *Proceedings of the Aristotelian Society, 114*, 103-21.

Berto, F., & Jago, M. (2019). Impossible worlds. Manuscript submitted for publication.

Bjerring, J.C. (2014). On counterpossibles. *Philosophical Studies, 168*, 327-353.

Bjerring, J. C. K. (2010). Non-ideal epistemic spaces (Doctoral dissertation). Retrieved from https://www.academia.edu/1020046/Non-Ideal_Epistemic_Spaces

Blackburn, P., Rijke, M., & Venema, Y. (2001). Modal logic: Cambridge tracts in theoretical computer science. Cambridge: Cambridge University Press.

Blackburn P., van Benthem J., Wolter F. (2006). Handbook of modal Logic, *Elsevier*.

Brogaard, B., & Salerno, J. (2013). Remarks on counterpossibles. *Synthese, 190*, 639-660.

Carnap, R. (1947). Meaning and necessity. Chicago: Chicago University Press.

Chalmers, D. (2008). Hyperintensionality and impossible worlds: An introduction. *Hyperintensionality and Impossible Worlds*. Presented at ANU School of Philosophy, Canberra.

Chang, C.C., Keisler, H.J. (2013). Model theory: Third edition. Courier Corporation.

Chellas, B. (1975). Basic conditional logic. *Journal of Philosophical Logic*, *4*, 133–228.

Cresswell, M. J. (1970). Classical intensional logics. *Theoria, 36*, 347–372.

Divers, J. (2002). Possible worlds. London: Routledge.

Dunn, J. M. (1976). Intuitive semantics for first-degree entailment and “coupled trees”. *Philosophical Studies, 29*, 149–168.

Field, H. (1989). Realism, mathematics and modality. Oxford: Oxford University Press.

Fine, K. (1975). Critical notice of Lewis, Counterfactuals. *Mind, 8*, 451 - 458.

Fitting, M. (2015). Intensional logic. *The Stanford Encyclopedia of Philosophy*. Retreived July 19, 2015, from<https://plato.stanford.edu/archives/sum2015/entries/logic-intensional/>.

Frege, G. (1884). Grundlagen der Arithmetik. Breslau: Wilhelm Koebner.

Gabbay, Dov M. (1972). A general theory of the conditional in terms of a ternary operator. *Theoria, 38*, 97-104.

Girard, P., & Triplett M. A. (2018). Prioritised ceteris paribus logic for counterfactual reasoning. *Synthese, 195*, 1681-1703.

Goodman, N. (1983). Fact, fiction, and forecast: Fourth edition. Cambrige, MA: Harvard University Press.

Goodman, N. (1972). Problems and projects. Indianapolis: Bobbs-Merrill.

Hájek, A. (2014). Most counterfactuals are false. Retrieved June 12, 2014, from http://Philrsss.Anu.Edu.Au/People-Defaults/Alanh/Papers/Mcf.Pdf.

Hintikka, J. (1962). Knowledge and belief. Ithaca: Cornell University Press.

Hintikka, J. (1975). Impossible possible worlds vindicated. *Journal of Philosophical Logic*, *4*, 475 - 484.

Jago, M. (2009). Logical information and epistemic space. *Synthese, 167*, 327-341.

Jago, M., (2012). Constructing worlds. *Synthese*, *189*, 59-74.

Jago, M. (2013). The content of deduction. *Journal of Philosophical Logic*, *42*, 317-334.

Jech, T. (2004). Set theory: Third millennium edition, revised and expanded. Springer Monographs in Mathematics.

Kiourti, I. (2010). Real impossible worlds: the bounds of possibility, (Doctoral dissertation). Retreived from https://www.academia.edu/1262571/Real_Impossible_Worlds_The_Bounds_of_Possibility

Kripke, S. (1980). Naming and necessity. Cambridge, MA: Harvard University Press.

Lewis, D. (1968). Counterpart theory and quantified modal logic. *Journal of Philosophy*, *65*, 113–26.

Lewis, D. (1970). Anselm and actuality. *Noûs*, *4*, 175–88.

Lewis, D. (1971). Completeness and decidability of three logics of counterfactual conditionals. *Theoria*, *37*, 74-85.

Lewis, D. (1973). Counterfactuals. Oxford: Blackwell.

Lewis, D. (1979). Counterfactual dependence and time's arrow. *Noûs*, *13***,** 455-476.

Lewis, D. (1980). A subjectivist's guide to objective chance. In Richard C. Jeffrey (Ed.), *Studies in Inductive Logic and Probability* (pp.83-132). University of California Press.

Lewis, D. (1981). Ordering semantics and premise semantics for counterfactuals. *Journal of Philosophical Logic*, *10*, 217-234.

Lewis, D. (1986). On the plurality of worlds. Oxford: Blackwell.

Lycan, W. (1994). Modality and meaning. Dordrecht: Kluwer.

Mares, E. D. (1997). Who's afraid of impossible worlds? *Notre Dame Journal of Formal Logic*, 38, 516-526.

Mares, E. D. (2004). Relevant logic: A philosophical interpretation. Cambridge: Cambridge University Press.

Mendelson, E. (1997). Introduction to mathematical logic: Fourth edition. Boca Raton: Chapman and Hall.

Moschovakis, Y.N. (1994). Notes on Set Theory. New York: Springer-Verlag.

Naylor, M. (1986). A note on David Lewis' realism about possible worlds. *Analysis*, *46*, 28-9.

Nolan, D. (1997). Impossible worlds: A modest approach. *Notre Dame Journal of Formal Logic*, *38*, 535-572.

Nolan, D. (2014). Hyperintensional metaphysics. *Philosophical Studies*, *171*, 141-160.

Nute, D. (1980). Topics in conditional logic. Dordrecht: D. Reidel Publishing Company.

Omori, H., & Wansing, H. (2017). 40 years of FDE: An introductory overview. *Studia Logica , 105*,1021-1049.

Pacuit, E. (2017). Neighbourhood Semantics for Modal Logic. Springer International.

Parsons, T. (1980). Nonexistent objects. New Haven: Yale University Press.

Pietarinen, A. (1998). Impossible worlds and logical omniscience: A note on MacPherson's logic of belief. *The 10th White House Papers: Graduate Research in the Cognitive & Computing Sciences at Sussex*, p. 8-13.

Pollock, J. L. (1976). The 'possible worlds' analysis of counterfactuals. *Philosophical Studies*, *29*, 469 - 476.

Priest, G. (1997). Sylvan's box: A short story and ten morals. *Notre Dame Journal of Formal Logic*, *38*, 573-582.

Priest, G. (2005). Towards non-being: The logic and metaphysics of intentionality. Oxford: Oxford University Press.

Priest, G. (2008). From if to is: An introduction to non-classical logic. Second edition. Cambridge: Cambridge University Press.

Priest, G. (2018). Some new thoughts on conditionals. *Topoi* , *27*, 369-77.

Prior A. N. (1955). Formal logic. Oxford: Clarendon Press.

Quine, W. (1960). Word and object. Cambridge, MA: MIT Press.

Rantala, V. (1982a). Impossible worlds semantics and logical omniscience. *Acta Philosophica Fennica*, *35*, 18-24.

Rantala, V. (1982b). Quantified modal logic: Non-normal worlds and propositional attitudes. *Studia Logica*, 41, 41-65.

Read, S. (1995). Thinking about logic: An introduction to the philosophy of logic. Oxford: Oxford University Press.

Rescher, N., & Brandom, R. (1980). The logic of inconsistency. A study in non-standard possible worlds semantics and ontology. Oxford: Blackwell.

Restall, G. (1997). Ways things can't be. *Notre Dame Journal of Formal Logic*, *38*, 583-596.

Restall, G. (2000). An introduction to substructural logics. Routledge.

Routley, R. (1980). Exploring Meinong's jungle and beyond. Canberra: Research School of the Social Sciences.

Sillari, G. (2008). Quantified logic of awareness and impossible possible worlds. *The Review of Symbolic Logic*, *1*, 514-529.

Skyrms, B. (1976). Possible worlds, physics and metaphysics. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition, 30*, 323-332.

Speaks, J. (2018). Theories of meaning. *The Stanford Encyclopedia of Philosophy*. Retrieved December 28, 2018, from <https://plato.stanford.edu/archives/win2018/entries/meaning/>.

Stalnaker, Robert C. (1968). A theory of conditionals. In N. Rescher (Ed.), *Studies in Logical Theory: American Philosophical Quarterly Monographs 2* (pp. 98-112). Oxford: Blackwell.

Stalnaker, R. C., & Thomason, R. H. (1970). A semantic analysis of conditional logic. *Theoria*, *36*, 23-42.

Stalnaker, R.C. (1980). A Defense of conditional excluded middle. In: Harper W.L., Stalnaker R.C., Pearce G. (Eds.) Ifs. *The University of Western Ontario Series in Philosophy of Science: A Series of Books in Philosophy of Science, Methodology, Epistemology, Logic, History of Science, and Related Fields* (pp.87-104). Dordrecht: Springer.

Vander Laan, D. (1997). The ontology of impossible worlds. *Notre Dame Journal of Formal Logic*, *38*, 597-620.

Vander Laan, D. (2004). Counterpossibles and similarities. In Frank Jackson & Graham Priest (Eds.) *Lewisian themes: The philosophy of David K. Lewis* (pp. 258-275). Oxford UK: Clarendon Press.

Varzi, A. (1997). Inconsistency without contradiction. *Notre Dame Journal of Formal Logic*, *38*, 621-39.

Wansing, H. (1990). A general possible worlds framework for reasoning about knowledge and belief. *Studia Logica: An International Journal for Symbolic Logic, 49*, 523-539.

Weiss, Y. (2017). Semantics for counterpossibles. *Australasian Journal of Logic*, *14*, 383-407.

Yagisawa, T. (1988). Beyond possible worlds. *Philosophical Studies*, *53*, 175–204.

Zalta, E. N. (1988). Intensional logic and the metaphysics of intentionality. Cambridge: MIT Press.