Counterfactuals

Notes by R.J. Buehler
Based on *Counterfactuals* by David Lewis

May 6, 2014

Contents

Preface

What follows are my personal notes on David Lewis' *Counterfactuals*. Most of the ideas presented in this document are not my own, but rather Lewis' and should be treated accordingly. This text is not meant for reproduction or as a replacement for Lewis' book, but rather as a convenient reference and summary, suitable for use as lecture notes or a review and little more. For a complete presentation of the thoughts and arguments presented, please see the full text of *Counterfactuals*.

Chapter 1

An Analysis of Counterfactuals

1.1 Conditionals

1.1.1 Overview

Conditionals–that is, statements of the form '*If...*, *Then...*'–have long been a sore point for those who desire a logic that corresponds closely to natural language; to make sense of the difficulties which have plagued this effort, we consider five classes of conditionals, ways that–traditionally–philosophers and logicians have thought it helpful to delineate the conditional construction. These classes are, by no means, definitive or mutually exclusive.

Definition: Material Conditional
The conditional of classical logic ($\rightarrow$); the material conditional is false only if the antecedent is true and the consequent is false. Otherwise, it is considered true. Note that the truth or falsity of the antecedent and consequent are judged independently of one another.

Example 1.
3 is a prime number $\rightarrow$ *the sky is blue*

Definition: Indicative Conditional
The standard conditional of natural language; indicative conditionals state that their consequent is true (in this world) if their antecedent is.

Example 2.
If Socrates is a man, then Socrates is mortal.

The material conditional of classical logic is, of course, intended to capture the usage of the indicative conditional of natural language–and, quite famously, seems to do a rather poor job. In particular, the so-called paradoxes of the material implication (e.g., $\neg p \rightarrow (p \rightarrow q)$) appear to provide good reason to suppose that the material conditional is not adequate (although Grice has argued against this conclusion with some success). C.I. Lewis, in an attempt to rectify the perceived deficiency of the material conditional, proposed the strict conditional as the proper logical representation for the indicative conditional:

Definition: Strict Conditional
Using the necessity operator of modal logic ($\Box$), a strict conditional is defined as

$$\Box(\phi \to \psi)$$

where ϕ and ψ are well-formed formulas of the logic and $\to$ is the material conditional. In terms of its intended interpretation, the strict conditional asserts that it is necessarily the case that $\phi \to \psi$.

Example 3.
$\Box$ *(Socrates is immortal → 2+2 =4)*

As the example given shows, the strict conditional–while escaping the paradoxes of material implication–still appears to struggle with proposition that are necessarily true or necessarily false.

Definition: Subjunctive Conditional
At its most basic, a subjunctive conditional is simply a natural language conditional using the subjunctive verb mood; the subjunctive mood is usually associated with states of unreality, with sentences expressing something which runs counter to what is actually the case.

Example 4.
If Socrates were immortal, he would not be human.
If I were to study all day tomorrow, I would get a perfect score on the quiz.

Definition: Counterfactual Conditional
A counterfactual conditional is a natural language conditional which asserts that its consequent would obtain if its antecedent were an accurate description of reality. Note that, following Lewis, this allows for counterfactual conditionals which are not, actually, counterfactual.

Example 5.
If Socrates were immortal, he would not be human.
If I were to study all day tomorrow, I would get a perfect score on the quiz.

A natural and important question is whether the distinctions given above are meaningful, especially those describing natural language. It is commonly accepted that there is a meaningful difference between indicative and subjunctive/counterfactual conditionals. The standard argument for this claim is owed to Ernest Adams (3):

Consider the sentences below:

If Oswald did not kill Kennedy, someone else did.
If Oswald hadn't killed Kennedy, someone else would have.

The first is an indicative conditional while the second is a subjunctive/counterfactual conditional. If there were no meaningful difference between the classes, two conditionals with only tense and mood differences as here should always receive the same truth value. Nevertheless, it is easy to imagine the first sentence being true and the second false. It follows, then, that there is a logical difference between indicative and subjunctive/counterfactual conditionals.

I don't know if I'm entirely on board with even the counterfactual-indicative distinction sketched so far; as Lewis points out, there seem to be conditionals of each type which behave more like the other (e.g., future subjunctives). Could indicative conditionals just be a special case of counterfactual conditionals (e.g., when the actual world matches the antecedent)?

Lewis' distinction between counterfactual and subjunctive conditionals has not, however, been as well received with most philosophers using the terms interchangeably. Lewis argues for the distinction by pointing out that future subjunctive conditionals seem to behave as if they were indicative conditionals and that a subjunctive verb mood isn't necessary to express a counterfactual conditional, e.g. 'If no Hitler, then no A-bomb'.

1.1.2 The Necessity of Being Vague

It's important to note that counterfactuals have, perhaps necessarily so, a certain vagueness; Lewis accepts this and argues that any explication of counterfactuals must therefore either be stated in vague terms or be made relative to some parameter which is fixed only within some rough limits.

This is a very important point and bears thinking about; at the moment, I tend to agree though.

1.2 Counterfactual Conditional Operators

1.2.1 The New Operators

We add two new operators '$\Box\!\!\rightarrow$' and ' $\Diamond\!\!\rightarrow$'. Taking each it turn,

$$\phi \Box\!\!\rightarrow \psi$$

for well-formed formulas ϕ and ψ, is meant to be read as *'If it were the case that ϕ, then it would be the case that ψ'*. Similarly,

$$\phi \Diamond\!\!\rightarrow \psi$$

for well-formed formulas ϕ and ψ, is meant to be read as *'If it were the case that ϕ, then it might be the case that ψ'*. The two counterfactual conditionals given are to be interdefinable in as follows:

$$\begin{array}{lll} \phi \Diamond\!\!\rightarrow \psi & :- & \neg(\phi \Box\!\!\rightarrow \neg\psi) \\ \psi \Box\!\!\rightarrow \psi & :- & \neg(\phi \Diamond\!\!\rightarrow \neg\psi) \end{array}$$

and so only one need be taken as primitive. Note further that the readings given above don't reflect the tense changes that sometimes occur with counterfactuals, nor do they imply that $\neg\phi$, as some are tempted to say natural language counterfactuals do.

1.2.2 Modal Logic

Modal logic makes use of a necessity operator ($\Box$) that operates as a kind of restricted universal quantification over *accessible* 'possible worlds' or states; something is said to be necessary if it is true at every world accessible from the base world. What worlds are accessible is, of course, varied with the notion of necessity under consideration. Similarly, modal logic also makes use of a possibility operator ($\Diamond$) which behaves as a restricted existential quantification over accessible worlds; some sentence ϕ is true if and only if there exists some world where ϕ is true which is accessible from the base world. As with necessity, what worlds are regarded as accessible is meant to be varied depending on the notion of possibility in play. It's worth noting that when the accessibility relation for a $\Diamond$ operator and a $\Box$ operator are the same, the two are interdefinable:

$$\begin{array}{lll} \Box\phi & :- & \neg\Diamond\neg\phi \\ \Diamond\phi & :- & \neg\Box\neg\phi \end{array}$$

In defining counterfactual conditional operators, we take a slightly non-standard, but equivalent formulation of the typical modal logic definition. Instead of an accessibility relation, a set of accessible worlds is assigned to each world for each necessity/possibility operator in use:

> **Definition: Sphere of Accessibility**
> The set of worlds accesible from a given world w for a given operator $\Box_i$; in symbols, s^i_w.

We extend the standard modal definitions of $\Box$ and $\Diamond$ in the obvious way:

- $\Box_i\phi$ is true at a world w if and only if for every $u \in s^i_w$, $u \models \phi$.
- $\Diamond_i\phi$ is true at a world w if and only if there exists $u \in s^i_w$ such that $u \models \phi$.

1.3 Fixed Strict Conditionals

1.3.1 Strictness Orderings

Considering the strict conditional $\Box_i(\phi \to \psi)$ presented earlier, shows that a partial ordering on the strictness of various conditionals exists. In particular, given any two non-identical necessity operators $\Box_i$ and $\Box_j$ and a world w, either:

- $s^i_w \subseteq s^j_w$
- $s^j_w \subseteq s^i_w$
- Neither $s^i_w \subseteq s^j_w$, nor $s^j_w \subseteq s^i_w$

in the first case, we say that $\Box_j(\phi \to \psi)$ is stricter at world w than $\Box_i(\phi \to \psi)$ and vice-versa for the second. In the last case, the two are incomparable. Extending this fixed-world partial ordering, we say that $\Box_j(\phi \to \psi)$ is stricter than $\Box_i(\phi \to \psi)$ if $s^i_w \subseteq s^j_w$ for all worlds w.

1.3.2 Counterfactuals as Strict Conditionals

Comparing strict conditionals in this manner gives rise to an intuitive conception of what a counterfactual conditional is: a strict conditional with a very particular sphere of accessibility; in particular that $\phi \Box\!\!\rightarrow \psi$ is true if and only if all worlds sufficiently similar to the base world make $\phi \to \psi$ true. Of course, 'sufficiently similar' is a very vague notion, but presume–for the moment–that we have some satisfactory way of balancing the changes between worlds and arriving at a similarity ordering for possible worlds. Lewis argues that, even in this idealized case, the proposed definition is flawed.

Consider the following:

> *If Jon had fired as soon as he saw Tom, Jon would have lived. If Tom had been wearing body armor, however, Jon would still not have lived.*

Translating these with ϕ as 'Jon had fired as soon as he saw Tom', ϕ' as 'Tom had been wearing body armor', and ψ as 'Jon would have lived', we have both:

$$\phi \Box\!\!\rightarrow \psi$$

$$(\phi \land \phi') \Box\!\!\rightarrow \neg\psi$$

Noting that the sentences above, in natural language, appear to be unproblematic together, our sketched analysis of counterfactuals should give the same result. Unfortunately, it does precisely the opposite. In order for the first counterfactual to be true, $\phi \to \psi$ must be true in every world sufficiently similar to the actual world; similarly, in order for the second to hold $(\phi \land \phi') \to \neg\psi$ must also be true in every one of these worlds as well. Yet, by the proposed definition, the first counterfactual asserts that every ϕ world under consideration is a ψ world. The second counterfactual, therefore, could only be true if no ϕ and ϕ' worlds were accessible; that is, it is vacuously true. In and of itself, this is very troubling since there seems to be little reason why such a world should a priori not exist. Matters, however, become even more desperate if we continue to add more to the antecedents of our counterfactuals, producing an entire sequence of counterfactuals which are all 'vacuously' true (by necessity), but appear to represent very real possible worlds.

Some deliberation shows that, at a basic level, the spheres of accessibility for counterfactuals–if they exist–are not static, that they likely change between our first and second sentences. The obvious solution to this is to consider the antecedent of the conditionals themselves as somehow determining the sphere of accessibility–a line of thought Lewis suggests is defeatist. Lewis suggests, rather, that counterfactuals are *variably strict conditionals*.

Is making the sphere of accessibility dependent on the antecedent really that bad? At least formally, it seems quite easy. For a base world w, the set of atomic sentences it makes true Γ, and counterfactual antecedent ϕ, take the set of worlds which make all or as many as possible of the atomic sentences in Γ true while making ϕ true; this is the set considered for the counterfactual evaluation. If we wanted to use this for reality, it would be better to use the world we believe ourselves to be in (not the actual world).

1.4 Variably Strict Conditionals

Definition: Variably Strict Conditional
Rather informally, a variable strict conditional is a strict conditional whose sphere of accessibility is not fixed, but rather varies as a result of–usually–the antecedent.

Definition: Centered System of Spheres
Let \$ be an assignment of each possible world w to a set of sets of possible worlds $\$_w$. \$ is called a *centered system of spheres* and the members of each $\$_w$ *spheres around* w if and only if, for every world w,

(C) $\$_w$ is *centered* on w; that is, $\{w\} \in \$_w$.

(1) $\$_w$ is *nested*; that is, for every $S, T \in \$_w$, either $S \subseteq T$ or $T \subseteq S$.

(2) $\$_w$ is *closed under unions*; that is, for every $\mathcal{S} \subseteq \$_w$, $\cup\mathcal{S} \in \$_w$.

(3) $\$_w$ is *closed under non-empty intersections*; that is, for every $\mathcal{S} \subseteq \$_w$ such that $\mathcal{S} \neq \emptyset$, $\cap\mathcal{S} \in \$_w$.

In practice, centered systems of spheres are often referred to as simply *systems of spheres*, unless it is ambiguous to do so.

Informally, a centered system of spheres gives a set of spheres representing varying degrees of similarity around a world w with the set $\{w\}$ representing the highest level of similarity and other spheres loosening the constraints from there. Since the centered system has nesting, the similarity ordering for two spheres of a given world is always well-defined. Lewis' conditions are, quite obviously, non-trivial, and he defends each in turn:

I'm not a large fan of this; it seems to require possible world omniscience of speakers which shouldn't be taken lightly

Also quite strong and non-obvious

(C) The constraint C asserts that the base world is more similar to itself than it is to any other world; while Lewis does consider a case where another world exists which is qualitatively indiscernible from the base world, he argues that the base world is still more similar to itself because it is identical with the base world.

(1) The constraint 1 asserts that the similarity ordering is a total ordering; that is, no two possible worlds are incomparable in regards to their similarity to the base world.

(2) Constraint 2 actually adds (almost) nothing not already guaranteed by (1) so long as the number of spheres is finite; if the number is infinite, however, it guarantees that–since $\cup\mathcal{S}$ satisfies all the properties of a sphere–it is considered one. As an unintended consequence, (2) also requires that $\emptyset$ is a sphere for every set of spheres $\$_i$–a fact which ultimately has no effect on the models.

(3) Finally, constraint 3 also only matters with infinite sets of spheres and asserts that–since $\cap\mathcal{S}$ satisfies all the properties of a sphere–it is considered one.

Definition: Universal
A centered system of spheres is called *universal* if and only if $\cup\$_w$ is the set of all possible worlds for every world w.

It's worth noting that if a centered system of spheres is not universal, any worlds not in $\cup\$_w$ for a world w are all considered equally similar to w and less similar than every world in $\cup\$_w$.

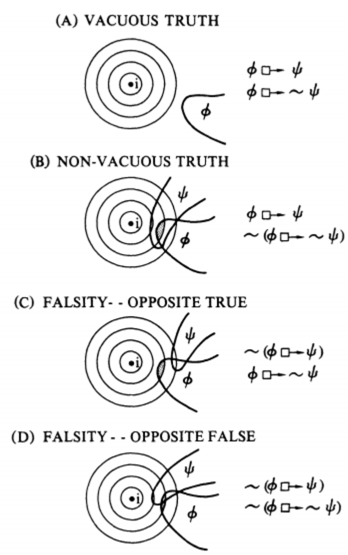
1.4.1 Truth Conditions

Definition: $\models$ '$\phi \mathrel{\Box\!\!\rightarrow} \psi$'
Alongside the standard truth conditions, we add the following: for a system of spheres $\$$ and world w,

$\$, w \models$ '$\phi \mathrel{\Box\!\!\rightarrow} \psi$' if and only if either

- $u \not\models \phi$ for every $u \in \cup\$_w$
- for some sphere $\mathcal{S} \in \$_w$, there exists $u \in \mathcal{S}$ such that $u \models \phi$ and $v \models$ '$\phi \rightarrow \psi$' for every $v \in \mathcal{S}$

Less formally, the first case accounts for when the counterfactual is vacuously true; in such cases, we will say that ϕ is *not entertainable*. The second case accounts for when the counterfactual is non-vacuously true; asking us to consider the smallest sphere which is *ϕ-permitting*–that is, in which there exists a ϕ world–and asserting that–within this sphere–the material conditional $\phi \rightarrow \psi$ is true. The figure below, taken from *Counterfactuals* gives diagrams for all possibilities:



Furthermore, some thought shows that the construction just given avoids the objection laid out against the fixed strict conditional formulation, allowing each conditional to generate its own sphere of accessibility, and thus failing to give the problematic result detailed there.

1.4.2 The Limit Assumption

Definition: The Limit Assumption
The assumption that for every world w and antecedent ϕ that is entertainable at w, there exists a smallest ϕ-permitting sphere; in other words, the inclusion ordering of spheres is a well-ordering.

As currently phrased, none of work requires or implies that we have the limit assumption; accepting it, however, does allow for a simplification of the truth conditions given earlier: $\$, w \models$ '$\phi \mathrel{\Box\!\!\rightarrow} \psi$' if and only if ψ holds at every ϕ world closest to w–note that no worlds outside of $\cup\$_w$ is considered. Unfortunately, constructing examples in which the limit assumption fails is relatively straightforward.

1.5 'Might' Counterfactuals and Outer Modalities

1.5.1 Truth Conditions

Recalling that, as proposed earlier, the 'would' and 'might' counterfactuals are interdefinable,

$$\begin{array}{lll} \phi \mathrel{\Diamond\!\!\rightarrow} \psi & :- & \neg(\phi \mathrel{\Box\!\!\rightarrow} \neg\psi) \\ \psi \mathrel{\Box\!\!\rightarrow} \psi & :- & \neg(\phi \mathrel{\Diamond\!\!\rightarrow} \neg\psi) \end{array}$$

the truth conditions for $\mathrel{\Box\!\!\rightarrow}$ now give rise to a truth definition for $\mathrel{\Diamond\!\!\rightarrow}$.

Definition: $\models$ **'$\phi \mathrel{\Diamond\!\!\rightarrow} \psi$'**
Alongside the standard truth conditions, we add the following: for a system of spheres $\$$ and world w,

$\$, w \models$ '$\phi \mathrel{\Box\!\!\rightarrow} \psi$' if and only if

- $u \models \phi$ for some $u \in \cup\$_w$
- for every sphere $\mathcal{S} \in \$_w$, if there exists $u \in \mathcal{S}$ such that $u \models \phi$, then there exists $v \in \mathcal{S}$ such that $v \models \phi$ and $v \models \psi$.

Just as with $\mathrel{\Box\!\!\rightarrow}$, these conditions could be simplified with the limit assumption; in particular–if the limit assumption holds–$\$, w \models$ '$\phi \mathrel{\Box\!\!\rightarrow} \psi$' if and only if the consequent holds at some antecedent world closest to w.

1.5.2 Generating a Modal Logic

Adding the symbols $\mathrel{\Box\!\!\rightarrow}$ and $\mathrel{\Diamond\!\!\rightarrow}$, it turns out, is sufficient to generate a proper modal logic. Letting '$\top$' denote a sentence which is always true and '$\bot$' a sentence which is always false,

$$\begin{array}{lll} \Diamond\phi & :- & \phi \mathrel{\Diamond\!\!\rightarrow} \top \\ \Box\phi & :- & \neg\Diamond\neg\phi \end{array}$$

The definitions given are, of course, only one of many choices ($\Diamond\phi$:- $\phi \mathrel{\Diamond\!\!\rightarrow} \phi$, for example, working just as well). Nonetheless, they accord not only with the standard interpretation of $\Diamond$ as 'it is possible that' / 'it is entertainable that', $\Box$ as 'it is necessary that' / 'it must be the case that', and the interdefinability of $\Box$ and $\Diamond$, but also with the standard, formal truth definitions of the operators:

Definition: $\models$ **'$\Box\phi$', $\models$ '$\Diamond\phi$'**
For a system of spheres $\$$ and world w,

- $\$, w \models$ '$\Box\phi$' if and only if for all $u \in \cup\$_w, u \models \phi$.
- $\$, w \models$ '$\Diamond\phi$' if and only if for some $u \in \cup\$_w, u \models \phi$.

The $\Box$ and $\Diamond$ defined correspond, therefore, to the sphere of accessibility $\cup\$_w$ for world w–the outermost sphere of accessibility for w and therefore strongest sense of necessity and possibility.

Definition: Outer Modalities
The notions of necessity and possibility corresponding to the sphere of accessibility $\cup\$_i$; that is, the $\Box$ and $\Diamond$ defined above. Individually, the $\Box$ above will be referred to as *outer necessity* and the $\Diamond$ above will be referred to as *outer possibility*.

Note that if $\$$ is universal, the outer modalities can be said to correspond to logical necessity and possibility.

1.5.3 A New System

Given that the standard modal operators are definable from $\Box\!\!\rightarrow$ and $\Diamond\!\!\rightarrow$, it's natural to wonder whether the system presented thus far is actually distinct from all the mainstream modal logics; perhaps, despite our efforts, $\phi \Box\!\!\rightarrow \psi$ is simply identical with $\Box(\phi \rightarrow \psi)$ or some more complicated construction.

Fortunately, this is not the case. To prove that this is the case, we need simply note that any standard modal formula depends only on the outer modalities defined above, and so will remain constant no matter what happens to the interior spheres. The counterfactual operators $\Box\!\!\rightarrow$ and $\Diamond\!\!\rightarrow$, however, do not, as a quick survey of their truth conditions shows. It follows immediately that the counterfactual operators cannot be identical with any standard modal formula.

1.6 Impossible Antecedents

According to the truth conditions thus far adopted for $\Box\!\!\rightarrow$ and $\Diamond\!\!\rightarrow$, counterfactuals with unsatisfiable antecedents are vacuously true; this choice, however, is not obvious. In favor of it, it seems as though we may be using counterfactuals when we engage in at least some reductio-style arguments, and this will only be possible if such counterfactuals are considered true.

Example 6.
If p_k were the largest prime, then $p_0 p_1 \ldots p_k + 1$ is a number which is not divisible by any prime, but isn't itself prime.

In addition, by making unsatisfiable antecedents true, counterfactuals which express consequents which are logically entailed by their antecedent are always true.

Example 7.
If it were the case that Socrates was mortal and Socrates was not mortal, then Socrates would be mortal.

Against this stance, however, is the possibility that–perhaps–not all counterfactuals with unsatisfiable antecedents should be treat equally; maybe there is another distinction to be made, maybe these aren't even counterfactuals at all. Moreover, it seems odd to assert any counterfactual with an unsatisfiable antecedent is true:

Example 8.
If it were the case that Socrates was mortal and Socrates was not mortal, then the sky would be blue.

Of course, this oddity could simply be owed to Gricean pragmatics, not any falsity on the part of the counterfactual; we don't desire, after all, to assert that the negation of the counterfactual above is true. Nonetheless, the case is not conclusive, and so we define a stronger pair of counterfactual operators which do not allow for vacuous truth.

Definition: $\Box\!\!\Rightarrow$
A stronger 'would' counterfactual operator which doesn't allow for vacuous truth. For a system of spheres $\$$ and world w, $\$, w \models$ '$\phi \Box\!\!\Rightarrow \psi$' if and only if there exists $\mathcal{S} \in \$_w$ such that for some $u \in \mathcal{S}$, $u \models \phi$, and for every $v \in \mathcal{S}$, $v \models$ '$\phi \rightarrow \psi$'.

Preserving the interdefinability of 'would' and 'might' counterfactual operators, we also introduce a new 'might' operator:

Definition: $\Diamond\!\!\Rightarrow$
A weakened 'might' counterfactual operator which allows for vacuous truth when its antecedent is unsatisfiable. For a system of spheres $\$$ and world w, $\$, w \models$ '$\phi \Diamond\!\!\Rightarrow \psi$' if and only if for every $\mathcal{S} \in \$_w$ such that for some $u \in \mathcal{S}$, $u \models \phi$, there exists $v \in \mathcal{S}$ such that $v \models \phi$ and $v \models \psi$.

As a final note, it's possible to derive either set of counterfactual operators from the others as follows:

$$\begin{aligned} \phi \Diamond\!\!\Rightarrow \psi \quad &:- \quad (\phi \Diamond\!\!\rightarrow \phi) \rightarrow (\phi \Diamond\!\!\rightarrow \psi) \\ \phi \Box\!\!\rightarrow \psi \quad &:- \quad (\phi \Box\!\!\Rightarrow \phi) \rightarrow (\phi \Box\!\!\Rightarrow \psi) \end{aligned}$$

1.7 True Antecedents and the Inner Modalities

1.7.1 True Antecedents

In natural language, there appears to be a presumption that the antecedents of counterfactuals do not, in reality, obtain; our current truth definitions, however, allow precisely this scenario. Indeed, with our assumption that the system of spheres is centered, a counterfactual with a true antecedent reduces to simply a material conditional. That is, both of the following inferences are valid:

$$\frac{\phi \wedge \neg\psi}{\therefore \quad \neg(\phi \;\Box\!\!\rightarrow \psi)} \qquad\qquad \frac{\phi \wedge \psi}{\therefore \quad \phi \;\Box\!\!\rightarrow \psi}$$

And thus, so are:

$$\frac{\phi \;\Box\!\!\rightarrow \psi}{\therefore \quad \phi \rightarrow \psi} \qquad\qquad \frac{\begin{array}{l}\phi \;\Box\!\!\rightarrow \psi \\ \phi\end{array}}{\therefore \quad \psi}$$

Despite many people's tendency to believe that the antecedent of a counterfactual must not obtain, there are natural language example which–at least at first blush–support the inferences above.

Example 9.

Speaker 1: *If Steve had come, we definitely wouldn't have had any candy left.*
Speaker 2: *That's not true; Steve did come, and there's a bunch of suckers left.*

Speaker 1: *If Steve had come, we would have had some of his home-made ice cream available too.*
Speaker 2: *You're right; Steve showed up a bit late,but if you hurry, there's still some ice cream left in the other room.*

Of course, against the inferences above are many of the so-called paradoxes of the material conditional; it seems odd–Grice aside–to assert that $\phi \;\Box\!\!\rightarrow \psi$ simply because both ϕ and ψ happen to be true in the actual world.

Lewis, despite believing that his analysis is correct, does offer an olive branch to those with different intuitions. Noting that the two original inferences above are licensed by different aspects of the (C) restriction on centered systems of spheres, Lewis suggests that the first–a result of the constraint that no world is more similar to a world w than w is to itself–is irreproachable, but that the second–a result of assuming that no world could be as similar to w as w is to itself–might be mistaken. This retraction gives rise to an alternate conception of a system of spheres:

1.7.2 Weakly Centered Systems and the Inner Modalities

This is much better.

Definition: Weakly Centered System of Spheres
Let \$ be an assignment of each possible world w to a set of sets of possible worlds $\$_w$. \$ is called a *weakly centered system of spheres* and the members of each $\$_w$ *spheres around* w if and only if, for every world w,

(W) $\$_w$ is *weakly centered* on w; that is, $\$_w \neq \{\emptyset\}$ and for every $\mathcal{S} \in \$_w$, if $\mathcal{S} \neq \emptyset$, $w \in \mathcal{S}$.

(1) $\$_w$ is *nested*; that is, for every $S, T \in \$_w$, either $S \subseteq T$ or $T \subseteq S$.

(2) $\$_w$ is *closed under unions*; that is, for every $\mathcal{S} \subseteq \$_w$, $\cup\mathcal{S} \in \$_w$.

(3) $\$_w$ is *closed under non-empty intersections*; that is, for every $\mathcal{S} \subseteq \$_w$ such that $\mathcal{S} \neq \emptyset$, $\cap\mathcal{S} \in \$_w$.

Leaving our truth definitions untouched, but using a weakly centered system of spheres instead of a centered one, the inference from $\phi \wedge \neg\psi$ to $\neg(\phi \mathrel{\Box\!\!\rightarrow} \psi)$ is still valid, but its counterpart is not; if $\phi \wedge \psi$ is true, $\phi \mathrel{\Box\!\!\rightarrow} \psi$ may be either true or false. Moreover, the model now distinguishes between truth at the base world w and truth within the smallest non-empty sphere of accessibility, $\cap(\$_i - \{\emptyset\})$. To express this latter, we introduce a new pair of modalities definable from our counterfactual operators:

$$\begin{aligned} \boxdot\phi &:- \quad \top \mathrel{\Box\!\!\Rightarrow} \phi \\ \Diamonddot\phi &:- \quad \top \mathrel{\Diamond\!\!\Rightarrow} \phi \end{aligned}$$

Definition: Inner Modalities
The notions of necessity and possibility corresponding to the sphere of accessibility $\cap(\$_i - \{\emptyset\})$–the innermost non-empty sphere of accessibility. For a system of spheres \$ and world w,

- $\$, w \models$ '$\boxdot\phi$' if and only if there exists a non-empty $\mathcal{S} \in \$_w$ such that for all $u \in \mathcal{S}$, $u \models \phi$.
- $\$, w \models$ '$\Diamonddot\phi$' if and only if for every non-empty $\mathcal{S} \in \$_w$, there exists $u \in \mathcal{S}$ such that $u \models \phi$.

Note that, assuming there is an innermost sphere around a world w, the truth conditions given simplify to $\boxdot\phi$ is true when every world within this innermost sphere is a ϕ world and $\Diamonddot\phi$ is true only when some world within this innermost sphere is a ϕ world. Similarly, there exist a number of equivalent definitions of $\boxdot$ and $\Diamonddot$ in our language (the $\mathrel{\Box\!\!\Rightarrow}$ and $\mathrel{\Diamond\!\!\Rightarrow}$ could, for instance, be $\mathrel{\Box\!\!\rightarrow}$ and $\mathrel{\Diamond\!\!\rightarrow}$) given our current assumptions; later, however, these assumptions may be weakened or changed, and so we take the formulation that is the most resistant to these changes. Finally, since the innermost sphere of accessibility must be a subset of the outermost, we have that $\Box\phi$ implies $\boxdot\phi$ and $\Diamonddot\phi$ implies $\Diamond\phi$.

It's worth mentioning as well that the inner and outer modalities together are still insufficient to define the counterfactual operators; the truth value of the counterfactual operator can depend, after all, on a sphere that is neither the innermost, nor the outermost.

1.8 Counterfactual Fallacies

A number of common inference patterns with the material conditional and fixed strict conditionals fail with the variably strict conditionals proposed here.

1.8.1 Strengthening the Antecedent

Typically taken to be

$$\begin{array}{c} \phi \,\Box\!\!\rightarrow \neg\psi \\ \hline \therefore \quad (\phi \wedge \chi) \,\Box\!\!\rightarrow \psi) \end{array}$$

it was the failure of this inference which first motivated the move from fixed strict conditionals to variably strict conditionals (recall Jon and Tom's shootout). It turns out, in addition, that a modal form of weakening also fails with variably strict conditionals:

$$\begin{array}{c} \Box(\chi \rightarrow \phi) \\ \phi \,\Box\!\!\rightarrow \neg\psi \\ \hline \therefore \quad \chi \,\Box\!\!\rightarrow \psi \end{array}$$

Consider, for example, the following:

> **Example 10.**
> *If I had left before 6am, I would have arrived home before noon. If I had left at exactly 3am, however, I wouldn't have arrived home before noon since–right at that moment–a piano fell several hundred feet and smashed into my doorstep. Yet, it's necessary that if I had left at 3am, I left before 6am.*

1.8.2 Transitivity

Since we can infer $\Box(\phi \rightarrow \psi)$ from $\phi \,\Box\!\!\rightarrow \psi$, the failure of transitivity with respect to $\Box\!\!\rightarrow$ has already been established; that is,

$$\begin{array}{c} \chi \,\Box\!\!\rightarrow \phi \\ \phi \,\Box\!\!\rightarrow \psi \\ \hline \therefore \quad \chi \,\Box\!\!\rightarrow \psi \end{array}$$

is not valid. A natural language example owed to Stalnaker makes the point quite clear:

> **Example 11.**
> *If Hoover had been born a Russian, he would have been a communist. If Hoover were a communist, he would have been a traitor. Yet, it's not true that if Hoover had been born a Russian, he would have been a traitor.*

Some thought shows that, in all cases of transitivity failure, the formula $\phi \,\Diamond\!\!\rightarrow \neg\chi$ is true. A valid, transitivity-esque inference, then, is:

$$\begin{array}{c} \neg(\phi \,\Diamond\!\!\rightarrow \neg\chi) \\ \chi \,\Box\!\!\rightarrow \phi \\ \phi \,\Box\!\!\rightarrow \psi \\ \hline \therefore \quad \chi \,\Box\!\!\rightarrow \psi \end{array} \qquad \text{or, equivalently,} \qquad \begin{array}{c} \phi \,\Box\!\!\rightarrow \chi \\ \chi \,\Box\!\!\rightarrow \phi \\ \phi \,\Box\!\!\rightarrow \psi \\ \hline \therefore \quad \chi \,\Box\!\!\rightarrow \psi \end{array}$$

Indeed, from the validity of these, we may establish both the validity of the following special case of transitivity and the simplified version given to its right:

$$\begin{array}{l} \chi \mathrel{\Box\!\!\rightarrow} \chi \wedge \phi \\ \chi \wedge \phi \mathrel{\Box\!\!\rightarrow} \psi \\ \hline \therefore \quad \chi \mathrel{\Box\!\!\rightarrow} \psi \end{array} \qquad\qquad \begin{array}{l} \chi \mathrel{\Box\!\!\rightarrow} \phi \\ \phi \wedge \chi \mathrel{\Box\!\!\rightarrow} \psi \\ \hline \therefore \quad \chi \mathrel{\Box\!\!\rightarrow} \psi \end{array}$$

1.8.3 Contraposition

One or both of the following inferences,

$$\begin{array}{l} \phi \mathrel{\Box\!\!\rightarrow} \psi \\ \hline \therefore \quad \neg\psi \mathrel{\Box\!\!\rightarrow} \neg\phi \end{array} \qquad\qquad \begin{array}{l} \neg\psi \mathrel{\Box\!\!\rightarrow} \neg\phi \\ \hline \therefore \quad \phi \mathrel{\Box\!\!\rightarrow} \psi \end{array}$$

contraposition is also invalid with counterfactual conditionals.

Example 12.
If Jon had said he was going to the party, Sarah would have as well. But if Sarah hadn't said she was going to the party, Jon would have–he's been avoiding her all week.

Note, however, that while contraposition is not valid with counterfactuals, modus tollens still is:

$$\begin{array}{l} \phi \mathrel{\Box\!\!\rightarrow} \psi \\ \neg\psi \\ \hline \therefore \quad \neg\phi \end{array}$$

1.9 Quantifiers and Counterparts

1.9.1 Potentialities

If the winner hadn't cheated, he wouldn't have won.

The sentence above has two very different interpretations; on one reading, it is a counterfactual with an impossible consequent:

the winner did not cheat $\Box\!\!\rightarrow$ the winner did not win

It cannot be the case, after all, that the 'winner' does not win. Yet, this isn't properly what the sentence above is taken to express. Rather, what is meant is the *de re* reading that the person who actually won would not have done so if they hadn't cheated.

> **Definition: Potentiality**
> A counterfactual property; in general, we assume that potentialities have the form 'α is an x such that $\phi(x)$' where α is an object, x a variable, and $\phi(x)$ a counterfactual formula representing the property ascribed.

One manner of representing a potentiality logically, then, is as the sentence $\exists x(x = \alpha \wedge \phi(x))$. Of course, the interactions between quantifiers and the counterfactual operators have not yet been addressed; nevertheless, it's important for the correct resolution of both potentialities and other *prima facie* meaningful counterfactuals.

Example 13.
Any winner who would have won even if he/she hadn't bribed the judge is throwing away good money.

$$\forall x[W(x) \wedge B(x,j) \wedge (\neg B(x,j) \Box\!\!\rightarrow W(x)) \rightarrow T(x)]$$

There was at least one ancient ruler who would have conquered most of the world if they had gunpowder.

$$\exists x[A(x) \wedge (G(x) \Box\!\!\rightarrow C(x))]$$

1.9.2 Counterparts

A natural first attempt to give truth conditions for counterfactual formulas is to simply extend the truth conditions for counterfactual sentences, perhaps letting the object from the actual world tag along throughout. This immediately raises the question, however, of transworld identity; how are we to say that this object from the actual world is identical with such and such object in this possible world–especially when the two could be arbitrarily different? If we wish to save this approach, it seems necessary to either pose transworld identity as basic or give some criterion, perhaps based in similarity, for what is identical with what.

My initial suggestion may actually be able to deal with this nicely (take constants for all objects in the base world; having the same constants makes you more similar, leaving some out less. If your constants have different atomic properties, you're considered more different. Nowhere, however, is transworld identity a problem. Objects are either defined as the same as some object in the actual world or not); thinking about formal models (as earlier) may also provide a means of giving a non-trivial and non-problematic explication of transworld identity for Lewis' conception.

Lewis' solution, however, is to deny transworld identity altogether, to argue that objects in two possible worlds are never identical with one another, but rather can be *counterparts*.

> **Definition: Counterpart**
> The counterparts for an object at a possible world w are those things existing in w that resemble it closely enough in important respects of intrinsic quality and extrinsic relations, and that resemble it no less closely than do other things existing in w. Furthermore, an object is its own unique counterpart in its own world and objects which inhabit no possible world, but are common across all are also considered their own unique counterparts (e.g. numbers).

With the counterpart relation in mind, it's necessary to reformulate the truth conditions of our counterfactual operators:

Definition: $\models$ '$\phi(x) \mathrel{\Box\!\!\rightarrow} \psi(x)$'
For a system of spheres $\$$ and world w,
$\$, w \models$ '$\phi(x) \mathrel{\Box\!\!\rightarrow} \psi(x)$' if and only if either

- $u \not\models \phi(x_u)$ for every counterpart of x in u, x_u, and every $u \in \cup\$_w$
- for some sphere $\mathcal{S} \in \$_w$, there exists $u \in \mathcal{S}$ and a counterpart x_u from u such that $u \models \phi(x_u)$ and $v \models$ '$\phi(x_v) \rightarrow \psi(x_v)$' for every counterpart of x in v, x_v, and every $v \in \mathcal{S}$

Beyond simply interpreting sentences *de re* and utilizing a single counterpart relation, it may be necessary to allow multiple counterpart relations differing in stringency or aspects of comparison in order to resolve even more complicated natural language constructions–like those owed to Goodman below.

Example 14.
If New York City were in Georgia, New York City would be in the south.

If Georgia included New York City, not all of Georgia would be in the south.

Chapter 2

Reformulations

Within this section we explore multiple reformulations of the original centered system of spheres; some of these re-castings will be equivalent to this original while others will represent only special cases.

2.1 Multiple Modalities

A multiple modalities approach to counterfactuals allows the representation of the counterfactual operator as a combination of modal operators and truth-functional connectives–assuming that only a finite number of sphere exist. In particular, we number the spheres around each world i, assigning S_i^1 to $\{i\}$, S_i^2 to the next largest, and so forth. If a particular world has fewer spheres than another, excess numbers are assigned to the largest sphere so that all worlds have a constant number of–not necessarily distinct–spheres. Associated with the nth sphere for each world is a pair of modal operators $\Box_n$ and $\Diamond_n$ representing the accessibility relation given by the spheres. It's not hard to see that the counterfactual operator can now be defined as follows:

$$\begin{aligned} \phi \;\Box\!\!\rightarrow \psi \quad &:- \quad (\Diamond_1\phi \wedge \Box_1(\phi \to \psi)) \vee \cdots \vee (\Diamond_n\phi \wedge \Box_n(\phi \to \psi)) \vee \neg\Diamond_n\phi \\ \phi \;\Diamond\!\!\rightarrow \psi \quad &:- \quad (\Diamond_1\phi \to \Diamond_1(\phi \wedge \psi)) \wedge \cdots \wedge (\Diamond_n\phi \to \Diamond_n(\phi \wedge \psi)) \wedge \Diamond_n\phi \end{aligned}$$

2.2 Propositional Quantification

I. Sets of Worlds as Propositions

   i. The proposition p is true at a world w if and only if $w \in p$

   ii. For a formula ϕ, the proposition expressed by ϕ is $[\![\phi]\!]$

      (a) All tautologies express the same proposition, namely the set of all worlds or *necessary proposition.*

      (b) All contradictions express the same proposition, namely the empty set or *impossible proposition.*

      Expressible A proposition is expressible in a language if and only if there is some sentence of the language which is true at all and only the worlds of the proposition.

   iii. The standard logical operators extend naturally into an algebra of propositions; if W is the set of all worlds,

$$\begin{aligned} [\![\neg\phi]\!] &= W - [\![\phi]\!] \\ [\![\phi \wedge \psi]\!] &= [\![\phi]\!] \cap [\![\psi]\!] \\ [\![\phi \vee \psi]\!] &= [\![\phi]\!] \cup [\![\psi]\!] \end{aligned}$$

iv. $[\![\phi \mathrel{\Box\!\!\rightarrow} \psi]\!]$ may similarly be cashed out, albeit with more complexity than those above

2.3 Comparative Similarity

What happens if we propose a Goodman-esque scenario wherein our counterfactual world is exactly the same as the actual world until some point in the future, at which point it goes haywire. Is it reasonable to say that this world is less similar than more, ostensibly, normal possible worlds? What if we don't know what the actual world is actually like? It seems the proposed world is one of the most similar, the most plausible!

I. The Similarity Relation

i. The system of spheres used earlier is simply a device to represent comparative similarity; to make this explicit,

> **Definition:** $\leq_w$
> For a given world w and two worlds i and j, $i \leq_w j$ if and only if i is more similar to w than j is

ii. The following six conditions are imposed on $\leq_w$ and S_w–the set of reachable worlds–for each w:

(a) $\leq_w$ is transitive
(b) $\leq_w$ is strongly connected (for any two worlds i, j either $i \leq_w j$ or $j \leq_w i$)
(c) $w \in S_w$
(d) w is strictly $\leq_w$-minimal (for all $i \in S_w$, if $i \neq w$, then $w <_w i$)
(e) Inaccessible worlds are $\leq_w$-maximal (if $k \notin S_w$, then $i \leq_w k$ for all worlds i)
(f) Accessible worlds are more similar to the base than inaccessible ones (if $k \notin S_w, i \in S_w$, then $i <_w k$)

II. Counterfactuals

i. The relation sketched above gives rise to a system exactly equivalent to the systems of spheres presented earlier; in particular,

> **Definition:** $\mathrel{\Box\!\!\rightarrow}$
> For sentences ϕ and ψ, '$\phi \mathrel{\Box\!\!\rightarrow} \psi$' is true at a world w according to a given comparative similarity system if and only if either

- no ϕ-world appears in S_w or
- there exists a ϕ world k in S_w such that, $\forall j \in S_w$, if $j \leq_w k$, then $j \models \phi \rightarrow \psi$

2.4 Comparative Possibility

Omitted; see pgs. 52-57

2.5 Cotenability

For the purposes of comparison latter on, we briefly sketch a cotenability-based formulation of the presented analysis:

> **Definition: Contenable**
> For sentences ξ, ϕ and world w, ξ is contenable with ϕ at w (according to a system of spheres $\$$) if and only if either

(1) ξ holds for all $w' \in \cup\$_w$

(2) ξ holds throughout some ϕ-permitting sphere in $\$_w$

Less formally, ξ is cotenable with ϕ if either ξ holds in every world accessible from w or there is a ϕ world closer to w than any $\neg\xi$ world. An outer necessity truth is therefore cotenable with everything while a falsehood is cotenable with nothing–other propositions lying somewhere in between.

> **Definition:** $\Box\!\!\rightarrow$
> For sentences ϕ and ψ, '$\phi \rightarrow \psi$' is true at a world w if and only if the premise ϕ and some auxiliary premise χ–cotenable with ϕ in w–logically imply ψ.

The definition above is provably equivalent to Lewis' original (pg. 57).

2.6 Selection Functions and Operators

Omitted; see pgs. 57-64

Chapter 3

Comparisons

As with any new theory, it's incumbent upon proponents of Lewis' system to show both how it accounts for those aspects of older theories which appear correct and improves upon those aspects of older theories which appear flawed.

3.1 Metalinguistic Theories

Most older theories of counterfactuals are *metaliguistic*, arguing for some variation on the theme that a counterfactual is assertable if and only if its antecedent–combined with suitable further premises–implies its consequent.

$$\phi \mathrel{\Box\!\!\rightarrow} \psi$$

is thus justified by an implicit, valid argument

$$\frac{\phi, \xi_1, \ldots, \xi_k}{\psi}$$

Less formally, counterfactual statements have been cast as both a claim that such a valid argument exists–a claim that is rather straightforwardly true or false–and as an elliptical statement of such an argument–a claim that cannot be judge true or false, but rather as valid or invalid when the argument is made explicit.

3.1.1 Implicit Premises

The looming challenge for metaliguistic theories, then, is to specify what kind of premises are suitable to use with a given counterfactual. In metalinguistic theories, these premises are used in much the same way as Lewis' system of spheres, ruling out the comparatively less similar possibilities in order to make the asserted connection true. It's common, and perhaps even quite reasonable, to assume that many of these premises are implicit in the context of our use of a counterfactual, either filling out the explicit antecedent of the counterfactual to its full, implicit form, resolving the vagueness of the comparison between competing counterfactual possibilities, or both. Consider Quine's pair of counterfactuals as an example:

Example 15.
If Caesar had been in command in Korea, he would have used the A-bomb.

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

3.1.2 Factual Premises

The preceeding is not, however, even theoretically sufficient; many counterfactuals are contingent and–even when made completely explicit–depend on the way the world actually is in order to obtain a truth value.

Example 16.
If I had looked in my pocket, I would have found a penny.

Furthermore, it seems highly suspect to claim that the fact of the matter in the example above is simply an implicit part of the original counterfactual claim; it simply doesn't seem to be the right sort of thing to be implicit in a conversation. It appears, then, that a metalinguistic theory must wield not only implicit premises to justify counterfactuals, but also factual ones. What relation a true statement must have to the explicit antecedent of the counterfactual is, of course, highly non-trivial; for the moment, we use Goodman's *cotenable* to describe a factual premise that has the right sort of relationship with the counterfactual antecedent.

So long as the nature of the cotenable relation remains mysterious, so does the metalinguistic account. Lewis himself proposed–as we saw in the last chapter–a definition for cotenable relying on his comparative similarity relation which makes the metalinguistic approach identical with his–a definition which is likely to curry little favor with metalinguists because of its reliance on possible worlds and comparative similarity. A more traditionally metalinguistic account is to hold that it isn't necessary to describe when a factual premise is cotenable with a counterfactual antecedent, but rather when it is *thought* to be cotenable with a counterfactual antecedent.

As Lewis points out, however, this method is more dangerous that it originally appears. The typical means of cashing out the approach above is to hold that a factual premise is cotenable with the counterfactual antecedent if we would still hold it to be true even if we came to know the counterfactual antecedent for certain–which gives precisely the wrong result for many counterfactuals. Consider,

Example 17.
If Oswald had not killed Kennedy, someone else would have.

If Oswald had not killed him, Kennedy would have lived.

Consider someone who holds it to be only marginally possible that Oswald did not assassinate Kennedy singlehandedly, but highly unlikely that Kennedy is still alive. Updating with the fact that 'Oswald did not kill Kennedy' would push them to assert the first counterfactual and not the second–the very opposite of intuition.

3.1.3 Laws of Nature

Omitted because this approach makes no sense to me; see page 72 of *Counterfactuals* if interested.

3.2 Stalnaker's Theory

The theory that comes closest to that presented by Lewis isn't any of the metaliguistic theories, but rather the approach developed by Robert Stalnaker. According to Stalnaker, a counterfactual $\phi \mathrel{\Box\!\!\rightarrow} \psi$ (he uses the notation $\phi > \psi$) is true if and only if either (1) ϕ is true at no world accessible from the actual world w or (2) ψ is true at the world closest to the base world w which makes ϕ true. Throughout his theory, Stalnaker relies not only on the limit assumption, but also that the similarity ordering on worlds is total–there are never two worlds which are equally similar to the base world. Note that, combined, this makes the comparative similarity ordering a well-ordering.

More formally, Stalnaker's system consists of an accessibility relation, a selection function f from counterfactual antecedents and base worlds to the nearest world accessible from the base world which makes the antecedent true, and an absurd world which makes everything true and so appears as the result of the selection function whenever no other world can. In order to interpret f as making selections based on comparative similarity, it's further required that $f(\phi, i) = i$ whenever i is a ϕ world and whenever ψ holds at $f(\phi, i)$ and ϕ holds at $f(\psi, i)$, $f(\phi, i) = f(\psi, i)$. Rather intuitively, then, Stalnaker's theory is a special case of Lewis'; indeed, Lewis proves as much.

3.2.1 The Counterfactual Excluded Middle

As Lewis so wonderfully puts it, both the principle vice and virtue of Stalnaker's system is the validity of the principle dubbed 'counterfactual excluded middle':

$$(\phi \mathrel{\Box\!\!\rightarrow} \psi) \vee (\phi \mathrel{\Box\!\!\rightarrow} \neg\psi)$$

In favor of such a principle is the fact that natural language doesn't seem to distinguish between the negation of a counterfactual $\neg(\phi \mathrel{\Box\!\!\rightarrow} \psi)$ and the negation of the consequent $\phi \mathrel{\Box\!\!\rightarrow} \neg\psi$. While both Lewis and Stalnaker's systems have the latter implying the former, only Stalnaker has the former implying the later and thus the validity of the principle of counterfactual excluded middle.

I. Lewis Against the Counterfactual Excluded Middle

- (i) Some natural language sentences seem like they should be true, but aren't with counterfactual excluded middle:
  - (a) *It is not the case that if Bizet and Verdi were compatriots, Bizet would be Italian; and it is not the case that if Bizet and Verdi were compatriots, Bizet would not be Italian; nevertheless, if Bizet and Verdi were compatriots, Bizet either would or would not be Italian.*
- (ii) Only in the vacuous case does $\Diamond\!\!\rightarrow$ differ from $\Box\!\!\rightarrow$ with Lewis' definitions
  - (a) The most plausible alternatives–$\Diamond(\phi \wedge \psi)$, $\Diamond(\phi \mathrel{\Box\!\!\rightarrow} \psi)$, $\phi \mathrel{\Box\!\!\rightarrow} \Diamond\psi$, and $\phi \mathrel{\Box\!\!\rightarrow} \Diamond(\phi \wedge \psi)$–also come out as unsatisfactory.
    - (1) Consider *'If I had looked in my pocket, I might have found a penny' when–in actuality–I did not look and no penny was there to be found.*
  - (b) Swapping possibility operators doesn't seem to help either

II. Stalnaker Revised

- (i) For those who desire a revised account of Stalnaker's theory which allows for ties in similarity between possible worlds (i.e. who find objection I.(i) above persuasive), but preserves the limit assumption, we may simply envision multiple selection functions available with no single function determined by the comparative similarity between worlds; call those selection functions which always give one of the closest worlds for a given world w and formula ϕ *admissible.*
- (ii) True, False, and Arbitrary/Neuter
  - A sentence is true at a world if it is true relative to all admissible selection functions
  - A sentence is false at a world if it is false relative to all admissible selection functions
  - A sentence is arbitrary/neuter at a world if it is not false or true relative to all admissible selection functions
- (iii) The revised theory validates all the same sentence schema as the original (including counterfactual excluded middle).
- (iv) Still seems to require $\Box\!\!\rightarrow$ and $\Diamond\!\!\rightarrow$ to be the same in non-vacuous cases

Chapter 4

Foundations

Much of the preceding work has progressed on two assumptions that, for varying reasons, have been called into question by various philosophers; we defend them here.

4.1 Possible Worlds

I. Defending Modal Realism

  i. **Modal Realism** Broadly construed, the stance the possible worlds are actual things, that there are possible worlds other than the one we inhabit that are just as real as it is.

  ii. Lewis' Argument from Natural Language

    (a) It's uncontroversial that things could have been different in countless ways–which is simply to say that there are many ways that things could have been besides the way they are. Each of these ways is a 'possible world'. Taking the reformulated statement at face value, it asserts the existence of possible worlds.

    (b) Of course, we don't take every statement at face value; nevertheless, there is a presumption to do so unless it causes trouble and there is an alternative which does not. No satisfactory argument exists for why modal realism is troublesome and no non-troublesome alternative exists.

  iii. Alternatives

    (a) Possible Worlds as Unanalyzed Primitives

      (1) This isn't an alternative; it's giving up.

    (b) 'Possibly ϕ' simply means that ϕ is a consistent sentence

      (1) If a consistent sentence is something like 'a sentence which could be true' or 'a sentence which isn't necessarily false' then the explanation is circular.

      (2) Leveraging Gödel's first incompleteness theorem, if a consistent sentence is one whose negation isn't provable in some specific formal system, then either there a falsehoods of arithmetic amongst the system's theorems or there is some truth of arithmetic not amongst its theorems.

      (3) If a consistent sentence is one which comes out true under the assignment of some possible extension to the non-logical vocabulary of the sentence, then the explanation is circular.

    (c) Possible Worlds are some respectable linguistic/mathematical entity

      (1) The leading options are 'maximal, consistent sets of sentences of some language', 'maximal, consistent sets of atomic sentences' (state descriptions), and 'maximal,

consistent sets of sentences in the language extended with constants naming every object'.

(2) All leverage a notion of consistency and depending on whether that notion is cashed out in modal or deductive/model theoretic terms, it is either circular or incorrect.

II. Arguments Against Modal Realism

  i. Realism is wrong because only our world actually exists

    (a) It's true that only our world *actually* exists because actually is an indexical which locates the statement as made in our possible world; realism is precisely the thesis that there is more than what *actually* exists.

  ii. Inflated Ontology

    (a) Theories can be parsimonious in two distinct ways: qualitatively and quantitatively. The first minimizes the number of kinds of things while the latter minimizes the number of instantiations of the kinds; modal realism is qualitatively parsimonious, but not quantitatively–which is the standard philosophy should be held to.

Is Lewis really any better than someone who takes possible worlds as basic?

4.2 Similarity

I. Ill-Understood or Vague

  i. A common objection to the approach outlined here is that it relies upon similarity between possible worlds which is a hopelessly unclear notion.

  ii. Lewis replies by distinguishing ill-understood notions (bad primitives) from vague ones (fine primitives). Counterfactuals are vague–why shouldn't one of our primitives be too?

  iii. Whatever we do with counterfactuals, comparative similarity is a common, vague notion

    (a) '*Seattle is more similar to Boston, than San Francisco*' is true or false depending upon what we judge as important–nonetheless, it is the type of assertion we make on a daily basis and must be dealt with, whatever we decide in regards to counterfactuals.

    (b) If counterfactuals and comparative similarity seem to vary together, we might as well have one vague notion rather than two.

Chapter 5

Analogies

Having presented not only an analysis of counterfactuals, but also multiple variations on this presentation, Lewis next gives three extensions of his work to areas that he argues use counterfactual-esque operators.

5.1 Conditional Obligation

I. Conditional Obligation

   i. Conditional obligation operators stem from the fact that sentences of the form 'If James robs the bank, then he should confess and return the loot' don't seem to have an unproblematic translation in terms of only $\boxdot$ and $\rightarrow$.

      (a) Note that the negation of the antecedent is obligatory, and so the consequent is as well (James shouldn't have anything to give back!). But, the consequent is obligatory given that the antecedent occurs.

   ii. To further justify a variably strict analysis, we may construct an alternating sequence as we originally did to justify a variably strict approach to counterfactuals:

      (a) *If James robbed the bank, he should confess*

      (b) *If James robbed the bank and confessing would send his mother to an early grave, he shouldn't confess*

      (c) *If James robbed the bank and confessing would send his mother to an early grave and an innocent man is on trial for his crime, he should confess*

II. Comparative Goodness

   i. Rather than a standard of comparative similarity, we may take a system of sphere as capturing some notion of comparative goodness.

   ii. Old Constraints

      (a) In such a setting, some of the basic assumptions of the counterfactual construction no longer seem tenable; in particular, centering and weak centering are counter-intuitive constraints.

      (b) As usual, the limit assumption seems dubious

      (c) Stalnaker's assumption seems even less likely in this context because not only does it seem that some good and some bad can balance each other, there also appear to be differences which have no bearing on comparative goodness (ties by irrelevance).

   iii. New Constraints

   (a) Normality - every world in the system of spheres must have at least one non-empty sphere in $\$_i$
      (1) Not enforcing this constraint allows worlds where nothing is obligatory, and everything is permissible.
   (b) Universality - for every $\$_i$ in $\$$, $\cup\$_i$ is the set of all possible worlds
      (1) This assumption blocks evaluability constraints; all worlds are evaluable in terms of comparative goodness from any world i
      (2) Note that universality implies normality
   (c) Absoluteness - $\$_i$ is the same for all i
      (1) This gives an objective scale of 'goodness' common across all the worlds

III. Operators

i. $\Box\!\!\Rightarrow$ and $\Box\!\!\rightarrow$
   (a) Carrying forward all the truth-functional definitions from the counterfactual analysis, $\Box\!\!\Rightarrow$ and $\Box\!\!\rightarrow$ can now be interpreted as the conditional obligation operator of deontic logic.
   (b) '$\phi \Box\!\!\rightarrow \psi$' and '$\phi \Box\!\!\Rightarrow \psi$' are therefore intended to be read as *given that ϕ, it is obligatory that ψ* or, even better, *given that ϕ, it ought to be that ψ*
   (c) If there are ϕ-worlds evaluable from i, then either conditional is true if and only if some $\phi \wedge \psi$-world is better, from the standpoint of i, than every $\phi \wedge \neg\psi$-world
   (d) Only in the case where ϕ is impossible, then, do '$\phi \Box\!\!\rightarrow \psi$' and '$\phi \Box\!\!\Rightarrow \psi$' differ–the former coming out vacuously true and the latter false.

ii. $\Diamond\!\!\Rightarrow$ and $\Diamond\!\!\rightarrow$
   (a) Carrying forward all the truth-functional definitions from the counterfactual analysis, $\Diamond\!\!\Rightarrow$ and $\Diamond\!\!\rightarrow$ can now be interpreted as the conditional permission operator of deontic logic.
   (b) '$\phi \Diamond\!\!\rightarrow \psi$' and '$\phi \Diamond\!\!\Rightarrow \psi$' are therefore intended to be read as *given that ϕ, it is permissible that ψ*
   (c) If there are ϕ-worlds evaluable from i, then either conditional is true if and only if some $\phi \wedge \psi$-world is at least as good, from the standpoint of i, as every $\phi \wedge \neg\psi$-world
   (d) Only in the case where ϕ is impossible, then, do '$\phi \Diamond\!\!\rightarrow \psi$' and '$\phi \Diamond\!\!\Rightarrow \psi$' differ–the former coming out false and the latter true.

iii. $\boxdot$ and $\diamonddot$
   (a) The inner modalities $\boxdot$ and $\diamonddot$ are versions of the standard O and P operators of deontic logic, deviating from the standard interpretation only when there is an infinite sequence of better and better worlds or the world itself is abnormal

iv. $\Box$ and $\Diamond$
   (a) The outer modalities $\Box$ and $\Diamond$ are the standard metaphysical necessity and possibility operators if we have universality; otherwise, they aren't familiar

IV. Comparison

i. For a comparison of Lewis' solution to van Fraassen and Hansson's see pages 102-104.

5.2 'When Next' and 'When Last'

I. Sentences depend for their truth on a host of different parameters; time, possible world, language/reference, all these factors come into play in assigning a truth value. Thus far, we've attempted to focus on only the possible worlds aspect by choosing examples that remain (mostly) fixed with respect to the others or for which there is a salient choice of a parameter. We consider now an application of the counterfactual analysis to variations in the temporal case.

   i. We will assume that time is linear, ignoring the possibilities of branches and loops

   ii. Letting $\leq$ and $<$ represent the linear order of time, we give the standard interpretations to the interval notation of mathematics, e.g. $[i, j)$ is all times k such that $i \leq k$ and $k < j$.

   iii. ∞ and $-\infty$ are allowed as a shorthand in the obvious way

   iv. Any intervals whose left coordinate is larger than its right is taken to be the empty interval.

II. Four Systems of Spheres

   i. The Future Temporal System of Spheres ($\$^F$)

      (a) Assigns to each time i the set $\$^F_i$ of all intervals beginning at i, but not including i itself (i.e., $(i, j], (i, j)$, and (i, ∞))

   ii. The Past Temporal System of Spheres ($\$^P$)

      (a) Assigns to each time i the set $\$^P_i$ of all intervals ending at i, but not including i itself (i.e., $(j, i), [j, i)$, and $(-\infty, i)$)

   iii. The Semi-Future Temporal System of Spheres ($\$^f$)

      (a) Assigns to each time i the set $\$^f_i$ of all intervals beginning at i and including i itself(i.e., $[i, j], [i, j)$, and $[i, \infty)$)

   iv. The Semi-Past Temporal System of Spheres ($\$^p$)

      (a) Assigns to each time i the set $\$^p_i$ of all intervals ending at i and including i itself (i.e., $(j, i], [j, i]$, and $(-\infty, i]$)

   v. Note that the semi-future and semi-past systems are centered while the future and past are not; none of the systems are necessarily normal. In the future or past systems, normality occurs if and only if there is no first/last time (these are the abnormal times).

   vi. Assuming the linearity of time makes Stalnaker's assumption and the limit assumption equivalent, but doesn't force any of the systems to meet either requirement.

III. Operators

   i. All of the counterfactual operators are defined as usual; we superscript the operators with $f, F, p,$ or P to indicate which system it derives from.

   ii. All of the operators may be read as varying understandings of 'when next that ϕ, it will be that ψ' or 'when last that ϕ, it was that ψ'

   iii. The difference between, say, the semi-future and future operators is simply in whether or not the present is considered to be available as a 'when next' state or not; additional differences between the operators arise in the vacuous case (ϕ is never true) and with infinitely many switches between $\phi \wedge \psi$ and $\phi \wedge \neg\psi$ worlds.

   iv. In the context of these systems, the outer modalities for the future and past systems are simply the standard temporal operators G, F, H, and P. The outer modalities for the semi-future and semi-past systems are the Diodorean temporal operators (essentially, each of the standard temporal operators, but also utilizing the present time, again in the obvious way).

   v. For a more extended discussion and, in particular, how Lewis' system compares against the work of Prior and Kamp see pgs. 108-111.

5.3 Contextually Definite Descriptions

I. Egocentric Logic

  i. In an interesting twist, we may think of sentences as being true at things (see Prior on Egocenteric logic for a more prolonged presentation).

  ii. We may think of sentences as possessing only a single variable which is interpreted as the object the sentence is being evaluated at. For example, 'x is a rock' is false at me, but true at any rock.

II. A Comparative Salience System

  i. Consider a system of spheres based on the comparative salience of objects; it is assumed throughout that a comparative salience ordering has at least the properties typical of a weak-ordering.

  ii. With our standard understanding of salience, such a system fails to be normal, centered, weakly-centered, absolute, and universal.

  iii. Defining the counterfactual operators in the normal way, the outer modalities become the existential and universal quantifiers over objects in the ken of the object under consideration.

III. The Variably Strict Egocentric Conditional

  i. The primary point of our analysis is, however, that the variably strict egocentric conditional '$\phi \Box\!\!\Rightarrow \psi$' can be understood as a 'contextually definite descriptions' operator. A 'contextually definite description' is taken to be a definite description ('the so-and-so') suitable for use when it is perfectly well-understand that there are many different so-and-so's, not just one.

    (a) The choice of $\Box\!\!\Rightarrow$ here follows from the obvious differences in the truth conditions of the various operators.

  ii. **Example 18.**
  "The pig is grunting" becomes 'x is a pig $\Box\!\!\Rightarrow$ x is grunting'

    (a) In terms of our truth conditions, the statement above is true if and only if there is some ϕ-permitting sphere (that is, a sphere which has a pig) such that ψ is true at all objects where ϕ is true (all the pigs are grunting).

  iii. As in our other cases, this analysis follows from the variably strict nature of contextually definite descriptions; note that, as earlier, we may construct an alternating sequence:

Example 19.

The pig is grunting.
The pig with spots is not grunting.
The Pig with spots and floppy ears is grunting.

Chapter 6

Logics

6.1 Completeness and Soundness Results

I. The Language

  i. Countably many sentence letters
  ii. The sentential constants $\bot$ and $\top$
  iii. The standard logical connectives $\neg$, $\vee$, $\wedge$, $\rightarrow$, and $\leftrightarrow$
  iv. The counterfactual operators $\boxright$, $\Box\!\!\Rightarrow$, $\Diamond\!\!\rightarrow$, $\Diamond\!\!\Rightarrow$, $\preceq$, $\prec$, $\approx$, $\Box$, $\Diamond$, $\boxdot$, and $\diamonddot$
  v. Punctuation $[,], (,)$
  vi. All of the symbols above generate the sentences of the language in the standard manner.

6.1.1 Model Theory

I. Interpretation

What's up with (8) and (9)?

Definition: Interpretation
An interpretation of the language based on a system of spheres $\$$ over a non-empty set I is a function $\llbracket\rrbracket$ from the sentences of the language to the power set of I such that for all sentences ϕ and ψ,

(1) $\llbracket\top\rrbracket = I$

(2) $\llbracket\bot\rrbracket = \emptyset$

(3) $\llbracket\neg\phi\rrbracket = I - \llbracket\phi\rrbracket$

(4) $\llbracket\phi \wedge \psi\rrbracket = \llbracket\phi\rrbracket \cap \llbracket\psi\rrbracket$

(5) $\llbracket\phi \vee \psi\rrbracket = \llbracket\phi\rrbracket \cup \llbracket\psi\rrbracket$

(6) $\llbracket\phi \rightarrow \psi\rrbracket = (I - \llbracket\phi\rrbracket) \cup \llbracket\psi\rrbracket$

(7) $\llbracket\phi \leftrightarrow \psi\rrbracket = (\llbracket\phi\rrbracket \cap \llbracket\psi\rrbracket) \cup ((I - \llbracket\phi\rrbracket) \cap (I - \llbracket\psi\rrbracket))$

(8) $\llbracket\phi \preceq \psi\rrbracket = \{i \in I : \forall S \in \$_i(\llbracket\psi\rrbracket \cdot S \rightarrow \llbracket\phi\rrbracket \cdot S)\}$

(9) $\llbracket\phi \prec \psi\rrbracket = \{i \in I : \exists S \in \$_i(\llbracket\phi\rrbracket \cdot S \wedge \neg\llbracket\psi\rrbracket \cdot S)\}$

(10) $\llbracket\phi \approx \psi\rrbracket = \{i \in I : \forall S \in \$_i(\llbracket\psi\rrbracket \cdot S \leftrightarrow \llbracket\phi\rrbracket \cdot S)\}$

(11) $\llbracket\Diamond\phi\rrbracket = \{i \in I : \llbracket\phi\rrbracket \cap \cup\$_i\}$

(12) $\llbracket\Box\phi\rrbracket = \{i \in I : \cup\$_i \subseteq \llbracket\phi\rrbracket\}$

(13) $\llbracket\diamonddot\phi\rrbracket = \{i \in I : \forall S \in \$_i(S \neq \emptyset \rightarrow \llbracket\phi\rrbracket \cdot S)\}$

(14) $\llbracket\boxdot\phi\rrbracket = \{i \in I : \exists S \in \$_i(S \neq \emptyset \wedge S \subseteq \llbracket\phi\rrbracket)\}$

(15) $\llbracket\phi \Box\!\!\Rightarrow \psi\rrbracket = \{i \in I : \exists S \in \$_i(\emptyset \neq \llbracket\phi\rrbracket \cap S \subseteq \llbracket\psi\rrbracket)\}$

(16) $\llbracket\phi \Diamond\!\!\Rightarrow \psi\rrbracket = \{i \in I : \forall S \in \$_i(\llbracket\phi\rrbracket \cdot S \rightarrow \llbracket\phi\rrbracket \cap \llbracket\psi\rrbracket \cdot S)\}$

(17) $\llbracket\phi \boxright \psi\rrbracket = \{i \in I : \llbracket\phi\rrbracket \cdot \cup\$_i \rightarrow \exists S \in \$_i(\emptyset \neq \llbracket\phi\rrbracket \cap S \subseteq \llbracket\psi\rrbracket)\}$

(18) $\llbracket\phi \Diamond\!\!\rightarrow \psi\rrbracket = \{i \in I : \llbracket\phi\rrbracket \cdot \cup\$_i \wedge \forall S \in \$_i(\llbracket\phi\rrbracket \cdot S \rightarrow \llbracket\phi\rrbracket \cap \llbracket\psi\rrbracket \cdot S)\}$

  i. where $A \cdot B$ is $A \cap B \neq \emptyset$ or, less formally, the boolean indicating whether A and B overlap
  ii. Definitions

Index Set- The set I for an interpretation

Indices- All $i \in I$ where I is the index set

Proposition- For a sentence ϕ, $[\![\phi]\!] \subseteq I$

True- A sentence ϕ is true at an index i if and only if $i \in [\![\phi]\!]$

Valid- A sentence ϕ is valid under an interpretation if and only if $I = [\![\phi]\!]$; a set of sentences is valid under an interpretation if and only if all its members are

iii. We require only that a system of spheres $\$_i$ is nested, closed under unions, and closed under non-empty intersections.

II. General Validity

> **Definition: General Validity**
> A sentence, set of sentences, or sentence schema is *generally valid* or *unconditionally valid* if and only if it is valid under every interpretation; a sentence, set of sentences, or sentence schema is *generally valid under conditions* or *conditionally valid* if and only if it is valid under every interpretation meeting certain conditions.

i. The conditions with which we are concerned are listed below:

(N)	$\$$ is *normal* if and only if for every $i \in I$, $\cup\$_i \neq \emptyset$
(T)	$\$$ is *totally reflexive* if and only if for every $i \in I$, $i \in \cup\$_i$
(W)	$\$$ is *weakly centered* if and only if for every $i \in I$ and every $S \in \$_i$ such that $S \neq \emptyset$, $i \in S$ and for at least one $S \in \$_i$, $S \neq \emptyset$
(C)	$\$$ is *centered* if and only if for every $i \in I$, $\{i\} \in \$_i$
(L)	$\$$ satisfies the *limit assumption* in relation to $[\![]\!]$ if and only if, for every ϕ, if $[\![\phi]\!]$ overlaps $\cup\$_i$, then there is some smallest member of $\$_i$ that overlaps $[\![\phi]\!]$.
(S)	$\$$ satisfies *Stalnaker's assumption* in relation to $[\![]\!]$ if and only if, for every ϕ, if $[\![\phi]\!]$ overlaps $\cup\$_i$, then there is some member of $\$_i$ whose intersection with $[\![\phi]\!]$ contains only a single index.
(U-)	$\$$ is *locally uniform* if and only if, for every $i \in I$ and $j \in \cup\$_i$, $\cup\$_i = \cup\$_j$
(U)	$\$$ is *uniform* if and only if, for every $i, j \in I$, $\cup\$_i = \cup\$_j$
(A-)	$\$$ is *locally absolute* if and only if, for every $i \in I$ and $j \in \cup\$_i$, $\$_i = \$_j$
(A)	$\$$ is *absolute* if and only if, for every $i, j \in I$, $\$_i = \$_j$
(UT)	$\$$ is *universal* if and only if, for every $i \in I$, $\cup\$_i = I$
(WA)	$\$$ is *weakly trivial* if and only if, for every $i \in I$, I is the only non-empty member of $\$_i$
(CA)	$\$$ is *trivial* if and only if $I = \{i\}$ and $\$_i$ contains both I and $\emptyset$

ii. We also have the following implications: $C \rightarrow W$, $W \rightarrow T$, $T \rightarrow N$, $S \rightarrow L$, $S \wedge W \rightarrow C$, $A \rightarrow A-$, $U \rightarrow U-$, $A \rightarrow U$, $CA \rightarrow WA$, $CA \rightarrow S$, $WA \rightarrow L$

III. Characteristic Axioms

(N)	Normality	$\top \prec \bot$
(T)	Totally Reflexive	$\Box\phi \to \phi$
(W)	Weakly Centered	$\Box\phi \lor \boxdot\phi \to \phi$
(C)	Centered	$\Diamond\!\!\!\cdot\,\phi \to \phi$
(L)	Limit Assumption	(None)
(S)	Stalnaker's Assumption	$\phi \land \psi \approx (\phi \land \neg\psi) \to \neg\Diamond\phi$
(U-)	Locally Uniform	$\Diamond\phi \to \Box\Diamond\phi$
(U)	Uniform	$\Box\phi \to \Box\Box\phi$
(A-)	Locally Absolute	$\phi \preceq \psi \to \Box(\phi \preceq \psi)$
(A)	Absolute	$\phi \prec \psi \to \Box(\phi \prec \psi)$
(UT)	Universal	U and T
(WA)	Weakly Trivial	W and A; or $\phi \preceq \psi := \Diamond\phi \to \Diamond\psi$
(CA)	Trivial	C and A; or $\phi \preceq \psi := \phi \to \psi$

i. Note that $U-/U$ and $A-/A$ above are inseparable in our language; adding the pair of schema as axioms gives both conditions. It follows that anything valid on the class of all $U-$ or $A-$ interpretations is still valid on the class of U or A interpretations.

ii. Taking the definitions given for $\preceq$ which produce interpretations meeting WA or CA respectively collapses our language into K for WA and standard propositional logic for CA.

6.1.2 Proof Theory

I. Variable Strictness Logics

i. Just as with the standard approach to modal logic, we specify a base logic, V, and obtain variations by adding the schema presented above as further axioms. For the sake of brevity, we specify V via an axiomatization:

Definition: V
The logic generated by the following inference rules:

- *Modus Ponens*
- Rule for Comparative Possibility: for any $n \geq 1$, if $\vdash \phi \to (\psi_1 \lor \cdots \lor \psi_n)$, then $\vdash (\psi_1 \preceq \phi) \lor \cdots \lor (\psi_n \preceq \phi)$

and the following axioms:

- All tautologies of propositional logic
- All definitions of operators
- Trans: $(\phi \preceq \psi) \land (\psi \preceq \xi) \to \phi \preceq \xi$
- Connex: $(\phi \preceq \psi) \lor (\psi \preceq \phi)$m

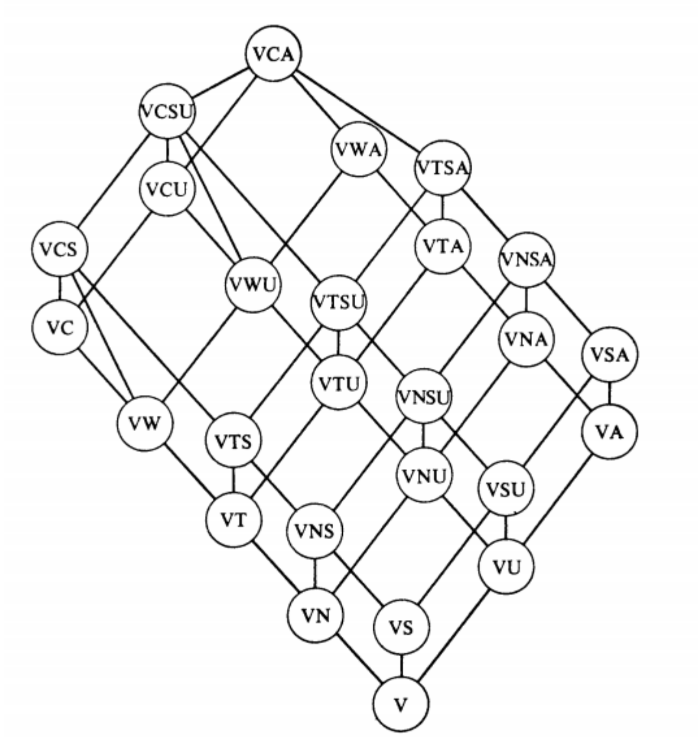
ii. The $\Box\!\!\to$ operator is derived from $\preceq$ as in §2.5 of *Counterfactuals*

iii. Connex and Trans together express that $\preceq$ is a weak ordering

Preserving truth versus preserving validity (pg. 123)?

It would be nice to have a natural deduction system for at least V...

II. An Overview



6.1.3 Completeness and Soundness

Omitted for the moment (see pgs. 124-131)

6.1.4 An Alternative Axiomatization of VC

Recalling Lewis' commentary during the analysis of counterfactuals, VC is Lewis' logic of choice for counterfactual analysis, and so he provides an axiomatization with $\Box\!\!\rightarrow$ taken as basic.

Definition: VC
The logic generated by the following inference rules:

- *Modus Ponens*
- Deduction within Conditionals: for any $n \geq 1$, if $\vdash \chi_1 \wedge \cdots \wedge \chi_n \to \psi$, then $\vdash (\phi \mathrel{\Box\!\!\rightarrow} \chi_1) \wedge \cdots \wedge (\phi \mathrel{\Box\!\!\rightarrow} \chi_n) \to (\phi \mathrel{\Box\!\!\rightarrow} \psi)$
- Interchange of logical equivalents

and the following axioms:

- All tautologies of propositional logic
- Definitions of non-primitive operators
- $\phi \mathrel{\Box\!\!\rightarrow} \phi$
- $(\neg\phi \mathrel{\Box\!\!\rightarrow} \phi) \to (\psi \mathrel{\Box\!\!\rightarrow} \phi)$
- $(\phi \mathrel{\Box\!\!\rightarrow} \neg\psi) \vee ((\phi \wedge \psi \mathrel{\Box\!\!\rightarrow} \chi) \leftrightarrow (\phi \mathrel{\Box\!\!\rightarrow} (\psi \to \chi)))$
- $(\phi \mathrel{\Box\!\!\rightarrow} \psi) \to (\phi \to \psi)$
- $\phi \wedge \psi \to (\phi \mathrel{\Box\!\!\rightarrow} \psi)$

Unfortunately, at least one long and unintuitive axiom seems necessary to produce VC–the fifth axiom in the axiomatization above.

6.2 Decidability Results

6.3 Derived Modal Logics

Chapter 7

Criticism and Reception

7.1 Fine 1975

I. Three Objections

   i. Nixon and the Bomb

      (a) 'If Nixon had pressed the button, then there would have been a nuclear holocaust.'

      (b) In natural language, this seems true. But surely–if it's similarity to the actual world under consideration–the closest possible world in which Nixon pressed the button is not a nuclear holocaust world! We may imagine any number of situations from a short power outage to a faulty button that make the antecedent of the counterfactual true, don't have a nuclear holocaust, and are eminently closer to the actual world than any nuclear holocaust world. Lewis' analysis therefore predicts that the sentence above is false–contra intuition.

      (c) More generally, Lewis cannot account for counterfactuals asserting that a small change leads to a large one; there will always be possible worlds which counteract the small change with another small change (and thus make the counterfactual false on his analysis).

   ii. Counterfactuals and the Actual World

      (a) True Antecedent, True Consequent

         (1) As presented, Lewis' analysis validates

$$\frac{\phi, \psi}{\phi \mathrel{\Box\!\!\rightarrow} \psi}$$

         which Lewis defends by claiming such counterfactuals are strange, but not–strictly speaking–false.

         (2) Consider the following case: My students have all taken their final exam, but I have yet to grade it. I assert, 'if student X worked hard, he would have passed'. In truth, however, while student X did work hard, he only passed because he cheated.

         (3) On Lewis' analysis, my assertion is true. Intuitively, it is false.

      (b) False Antecedent, True Consequent

         (1) If an irrelevant antecedent is false and the consequent true, the closest world always seems to be one that simply switches the antecedent.

         (2) 'If I lifted my pinky, Obama would have been reelected'

         (3) Fine argues that the rather the 'opposite' seems true: 'If I hadn't lifted my pinky, Obama would have been reelected'. Thinking about it, however, this seems entirely

explainable in terms of Gricean pragmatics (note that we may assert both of the preceding antecedents)

Bibliography

[1] Lewis, David. *Counterfactuals*