COUNTERFACTUALS AND CONDITIONAL PROBABILITIES

BY

ELANA SARA-LEAH TZIVIA GELLER

A Thesis
Submitted to the Faculty of Graduate Studies
In Partial Fulfillment of the Requirements for the Degree of

MASTER OF ARTS

Department of Philosophy
University of Manitoba
Winnipeg, Manitoba

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

*****

COPYRIGHT PERMISSION

COUNTERFACTUALS AND CONDITIONAL PROBABILITIES

BY

ELANA SARA-LEAH TZIVIA GELLER

A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of Manitoba in partial fulfillment of the requirement of the degree

of

MASTER OF ARTS

Contents

Abstract. iv
Acknowledgements v
Introduction 1
Chapter One: The Problem of Counterfactuals and Its Recent History 3
1. The Problem 3
2. History 4
3. Possible World Semantics 9
Chapter Two: Lewis, Possible World Semantics and Realism 12
1. Possible World Semantics 12
2. Lewis' Defense 18
3. Analysis 21
Chapter Three: Conditional Probability and Counterfactuals 27
1. Core Claim 28
2. The Stalnaker Thesis and Lewis' Triviality Claim 34
3. The *Mind* Debate 38
4. Possible World Semantics vs. Probabilistic Semantics 46
Bibliography 49

Abstract

The problem of counterfactuals arises due to the truth conditions for the material conditional. The truth conditions for the material conditional state that if a conditional has a false antecedent it is true. But there are many conditionals which we would like to maintain are false which are true according to these conditions. There are also true counterfactuals whose truth we would like to represent non-vacuously.

One of the most widely held solutions to the problem of counterfactuals is possible world semantics, proposed (most notably) by David Lewis. Lewis maintains that possible worlds are as real as this world. This seems unintuitive.

In this thesis, I forward a probabilistic semantics for counterfactuals. I maintain that the truth of a counterfactual equals the probability of its consequent given the truth of its antecedent. Due to some problems that arise I also state that the probability used ought to be interpreted on a frequentist account. I maintain that this semantics is better than a possible world account because the ontology is less suspect and one has epistemic access to crucial data, whereas, with possible worlds one does not.

Acknowledgements

I would like to thank my advisor, Bob Bright, without whom this would not have been possible. His support and guidance were invaluable. I would also like to thank my second reader, Kenneth Warmbrod, and my external reader Tom Kucera. Apart from my committee I would like to thank Jason Holt who spent hours talking with me about my thesis and gave me very useful suggestions and support.

Introduction

"If I were a rich man, Ya ha deedle deedle, bubba bubba deedle deedle dum"
from *Fiddler on the Roof*

The counterfactual conditional (also known as contrary-to-fact conditional) is an important part of language. A conditional is a sentence in the form of "If X, then Y." A counterfactual is a conditional which has a false X (i.e., the first part of the sentence, the antecedent, is false). It turns out that this causes many problems for the truth conditions. Finding truth conditions for the counterfactual is very important, because it is used so often. Although Chisholm found no (ultimate) solution to the problem of counterfactuals he indicated why they are so important to justify. The five reasons he gave for their importance are:

1) History and science are dependent on conditionals where the antecedent is known or believed to be false.
2) Precautionary activity is dependent on conditionals where the truth of the antecedent is not known.
3) Experiments depend on conditionals where one deliberately pretends that the antecedent is false.
4) Ordinary language is dependent on dispositions.
5) Many philosophic theories are dependent on counterfactuals.[1]

Thus, if we are to be justified in our moral claims, scientific inferences, philosophic theories, or common every day use of the counterfactual we must deal with the problem.

In this thesis I will argue for a solution to the problem of counterfactuals. In the first chapter I will explain the problem of counterfactuals in more detail and then briefly discuss the history of the dialogue concerning the problem. In the second chapter I will explicate the leading solution to the problem of counterfactuals, possible world

[1] Chisholm (1946), pp. 119-20.

semantics, most notably forwarded by David Lewis. In the third chapter I will argue that a theory based on conditional probability is a better account of counterfactuals than possible world semantics because although there are problems with regard to the interpretation of probability, there are no ontological problems of the sort that possible worlds raise.

Chapter One: The Problem of Counterfactuals and Its Recent History

1. The Problem

In elementary logic one is taught that the conditional (an "if_, then_" statement, p→q) is of one sort: the material conditional. The truth-conditions for the material conditional are: p→q is true iff either both p and q are true, or p is false; p→q is false iff p is true and q is false. These truth-conditions work well within the context of elementary logic but they do not work well for all that one wants to represent conditionally.

The counterfactual conditional is a conditional with a (typically) false antecedent, but a conditional which one does not want to automatically say is true, based only on the truth-conditions for the material conditional.[2] One example of a counterfactual which illustrates the problem is: If Galileo had leapt from the top of the tower of Pisa, then he would have turned into a bird and flown away.[3] One wants to say that this statement is false, even though it can be translated into sentential logic (SL) as a material conditional, in which case it would be true just because its antecedent is false.[4] It seems odd to say that the statement about Galileo is true just because he never jumped from the tower of Pisa. What one wants to say is that Galileo *would not have* turned into a bird, *even if he had* jumped off of the tower and, therefore, the first statement is false. The same problem arises with the true counterfactual conditional. If the above conditional were changed to

---

[2] "Typically" is used because there are counterfactuals with true antecedents. This will be discussed further when talking about Lewis' treatment of "would" and "could" conditionals.
[3] Warmbrod (1999), Chap. B.
[4] SL is a logical system which analyzes simple sentences joined by the connectives roughly equivalent to "or", "and", "not", "if _then" and "if and only if."

"If Galileo had leapt from the tower of Pisa, then he would have fallen to the ground" one would want to say that the conditional is true. The second Galileo-leaping-statement would be true (as was the first), according to the truth-conditions for sentential logic, but only trivially. The material conditional with a false antecedent is vacuously true (i.e., trivially true), yet, one wants more than vacuous truth for the second statement. One wants to be able to represent the connection between the antecedent and the consequent in these types of statements. If the logician relies on the truth-conditions for the material conditional the counterfactual would be true when it ought to be false, and only vacuously true when it ought to be meaningfully true.

The material conditional is truth-functional, i.e., one just has to see whether the two parts, on either side of the connective, are true or false to determine the truth of the whole sentence. With the material conditional one could test for the truth of "If it is raining outside, then John will come home" just by seeing if John comes home when it is raining, whether or not the rain has anything to do with John's arrival. As stated above, counterfactual conditionals are different, and their difference has to do with the special relationship *between* their antecedent and consequent.

2. History

There are two major approaches in solving the problem of counterfactual conditionals recognized in the literature: "the linguistic approach" and "the possible world approach."[5] The linguistic approach uses physical laws to explain counterfactuals. It is believed by its advocate that what connects the two sides of the counterfactual

[5] Pollock (1976), chap. 1.

conditional is not the truth of either or both sides but whether a certain physical law obtains. Someone who asserts a counterfactual according to this approach is asserting the existence of a valid argument which includes the antecedent as a premise and includes laws of nature and true statements about the world also as premises. This argument ends with the consequent as the conclusion. Thus, if I said "If I had thrown this glass on the floor, it would have broken", there would be some physical law that would explain this, i.e., gravity, molecular makeup of the glass, etc.

As Pollock points out, too many counterfactuals would turn out to be true if there were no restrictions on what truths might be included among the premises of an argument supporting a given counterfactual. He states:

> Without restrictions on what we can put into the circumstances C, it would turn out that whenever the material conditional "P⊃Q" is true, so is the subjunctive conditional "P>Q." This is because if "P⊃Q" is true, we can put it among the circumstances C, and then C together with P (without even making use of any laws) would automatically entail Q.[6]

Therefore, most of the literature on the linguistic approach is aimed at figuring out what these restrictions ought to be. Chisholm and Goodman made the two most notable attempts.

Chisholm tried the restriction of convention.[7] Whatever is commonly held to be the restriction is the right one. Chisholm said we can put any true sentence into our argument as relevant information. Thus, if my counterfactual were "if I struck this match it would light" the argument could include true sentences like "fire needs oxygen", "for a match to light it must be dry", etc. These true sentences as premises along with the antecedent will lead, according to Chisholm, to the truth of the consequent. Chisholm

[6] Ibid., p. 4.
[7] Chisholm (1946), p. 293.

abandoned this road of argument when he realised it leads to absurdities. He asked us to suppose that somebody accepts the following statements:

> (1) All gold is malleable;
> (2) No cast-iron is malleable;
> (3) Nothing is both gold and cast-iron;
> (4) Nothing is both malleable and not malleable;
> (5) That is cast-iron;
> (6) That is not gold; and,
> (7) That is not malleable.
> We may contrast three different situations in which he asserts three different Counterfactuals having the same antecedent. First, he asserts, pointing to an object his hearers don't know to be gold and don't know not to be gold, 'If that *were* gold, it would be malleable.' In this case, he is supposing the denial of (6); he is excluding from his presuppositions (5), (6) and (7); and he is concerned to emphasize (1). Secondly, he asserts, pointing to an object he and his hearers agree to be cast-iron, 'If *that* were gold, then some gold things would not be malleable.' He is again supposing the denial of (6); he is excluding (1) and (6), but he is no longer excluding (5) or (7); and he is concerned to emphasize either (5) or (2). Thirdly, he asserts, 'If that were gold, then some things would be both malleable and not malleable.' He is again supposing the denial of (6); he is now excluding (3) and no longer excluding (1), (5), (6) or (7); and he is now concerned to emphasize (1), (2) or (5).[8]

With this example Chisholm believes he shows that he cannot use convention to test the truth of counterfactuals. According to the linguistic analysis with the added constraint of convention one would have to be justified in asserting the third counterfactual. But this is problematic because one ought not be able to say that a thing has and does not have a trait such as malleability.

Goodman calls his restriction "cotenability."[9] Where A stands for the antecedent and S stands for a set of true sentences, Goodman defines cotenability as: "*A* is cotenable with *S* ... if it is not the case that S would not be true if A were."[10] This constrains the

[8] Chisholm (1955), p. 103.
[9] Goodman (1947), p. 120.
[10] Ibid.

addition of the set of sentences in that it cannot be the case the set of sentences would not be true where the antecedent was. With the same example it would mean that the set of sentences that are included in the argument like "the match is dry" and "there is oxygen in the room" cannot be false while "I lit the match" is true. What this constraint accomplishes is that one would not hold a counterfactual to be true where the conditions are impossible. This could be one of the problems with the conventionality restriction because it could allow a sentence that is false given the truth of the antecedent.

Goodman himself notices a problem with cotenability. He says:

> In order to determine the truth of a given counterfactual it seems that we have to determine, among other things, whether there is a suitable S that is cotenable with A and meets certain further requirements. But in order to determine whether or not a given S is cotenable with A, we have to determine whether or not the counterfactual "if A were true, then S would be true" is itself true... Thus we find ourselves involved in an infinite regressus...for cotenability is defined in terms of counterfactuals, yet the meaning of counterfactuals is defined in terms of cotenability...so that the problem of counterfactuals must remain unsolved.[11]

Since cotenability leads to a regress Goodman discards it.

Pollock states that the problem with the linguistic approach lies in the problem with physical law. He states that physical laws seem just as problematic as counterfactuals.[12] Pollock claims that physical laws, although considered by many to be derived from material conditional generalizations, are subjunctive in nature.[13] Thus, the problem with the use of physical law is much like the problem with cotenability, it results in a regress.

---

[11] Ibid., p. 121.
[12] For a discussion of physical laws and a thorough analysis cf. Goodman (1947), pp.123-8.
[13] Pollock (1976), p. 13.

Edgington, who will be spoken about in more detail in the last chapter, also has a criticism of the linguistic approach. She states that it is not likely that all counterfactuals which we think are true are governed by law, i.e., that their antecedents together with the laws of nature entail the consequent.[14]

Quine did not think there was a problem, *per se*, with counterfactuals. He said that the counter-to-fact conditional was just an idiom of speech. Quine thought dispositions were the "most well behaved" counterfactuals.[15] He stated that "the difference [between dispositions and ordinary counterfactuals] is that . . . a stabilizing factor is introduced [in the former]; a theory of subvisible structure."[16] He adds that when one speaks in counterfactuals terms she believes there is a subvisible structure which explains the knowledge the event x will happen to object o if event y obtains. In other words, Quine believes that there is a subvisible structure to those things that we call "fragile" which explains our calling them fragile. Thus, he believes there is no "magic" forethought, no hypothetical reasoning, there are just the properties in hand.

The problem with this solution is that it still rests on counterfactuals. No longer does one talk about what would happen to the object if event x took place, now one speaks about what would happen at the subvisible level if event x took place. So instead of explaining away the problem of counterfactuals Quine has just moved the worry from the visible to the nonvisible level.

---

[14] Edgington (1995), p. 249.
[15] A "disposition" is an object's tendency to act in a certain way when in a certain situation. For example, "fragility" is the disposition to break when hit or dropped. There is an ongoing debate as to whether dispositions ought to be analyzed in terms of counterfactuals.
[16] Quine (1960), p. 223.

He might respond that he has not moved the problem. Quine would insist that there is no counterfactual in the molecular structure of a thing. Counterfactuals are sentential. But if one were to ask Quine what it would *mean* for a thing to have a disposition at a subvisible level he would have to say that, in the case of fragility, the molecules were set up in such a way that if they were in such-and-such a situation the result would be they would break apart. This is still subjunctive. Thus the counterfactual problem is not solved.

3. Possible World Semantics

The concept of possible worlds has been discussed since the modern philosophers. Leibniz's famous line, "this, the best of all possible worlds", foreshadows possible world usage. He may not have been trying to solve the problem of counterfactuals but the concept of other worlds was alive none-the-less. Saul Kripke was the first to use possible worlds while formalising modal expressions. In his completeness theorems for different modal logics he used the concept and built a semantic analysis. Let K be the set of all possible worlds, R a relation on K (the "accessibility" relation), and ϕ a function assigning truth values to atomic propositions at each world. Kripke's truth conditions for modal statements are as follows:

> Given any two worlds $H_1, H_2 \in K$, we read "H_1 R H_2" as H_2 is "possible relative to H_1", "possible in H1" or "related to H_1"; that is to say, every proposition true in H_2 is to be possible in H_1... [W]e evaluate a formula *A* as *necessary* in a world H_1 if it is *true* in every world possible relative to H_1; i.e., $\phi(\sim A, H_1) = T$ iff $\phi(A, H_2) = T$ for each H_2 such that H_1 R H_2... *A* is possible relative to H_1 iff there exists H_2, possible relative to H_1, in which A is *true*.[17]

[17] Kripke (1963), p. 70.

Thus, a statement A is necessary if it is true in all worlds and possible if true in only some.

Stalnaker has a detailed explanation of how one is supposed to deal with counterfactuals. He states:

> First, add the antecedent (hypothetically) to your stock of beliefs; second, make whatever adjustments are required to maintain consistency (without modifying the hypothetical belief in the antecedent); finally, consider whether or not the consequent is then true.[18]

He further states that possible worlds are the testing tool, which is analogous to the stock of hypothetical beliefs. Stalnaker then explicates a model structure following Kripke:

> Let M be an ordered triple (K, R, λ). K is to be understood intuitively as the set of all possible worlds: R is the relation of relative possibility which defines the structure. If % and ! are possible worlds (members of K), then %R! reads "! is possible with respect to %." This means that, where % is the actual world, ! is a possible world. R is a reflexive relation; that is, every world is possible with respect to itself.[19]

(λ stands for an absurd world for impossible statements.) To test a counterfactual one selects the closest world to the actual world, one which does the least amount of "violence" to this reality.[20] The closest world would be that world that is exactly the same except for those minimal differences that are required to have the antecedent obtain. If in this world the consequent also obtains then the counterfactual is true.

In his seminal paper Lewis explains a few different alternative analyses of counterfactuals and then explains why they are problematic. He then goes on to explicate his own analysis of counterfactuals also following Kripke-like semantics.

---

[18] Stalnaker (1981a), p. 44.
[19] Ibid., p. 46.
[20] Ibid.

Lewis' analysis is like Stalnaker's except its uses spheres of nested worlds with a nearness relation: "Set S of worlds $=_{df}$ a *sphere* around i iff every S-world is accessible from i and is closer to i than is any non-S world."[21] No longer is there any one world which is closest to the actual world but groups of worlds which are closer than others.[22]

Lewis adds spheres to the possible world analysis because, he believes, Stalnaker's formulation of his nearness relation is problematic. He states that there may not be exactly *one* nearest world.[23] Another problem with Stalnaker's nearness relation is that it requires that there be a *closest* world.[24] If there are infinitely many worlds one can imagine there are infinitely many closer worlds. If I was looking to analyse the conditional "If I were any other height, then..." the nearest world may be the world in which I am only a millimetre taller or shorter than I am now. Putting aside the fact that there would be two different ways to go, i.e., the worlds where I am shorter and the worlds where I am taller (which illustrates why Lewis has misgivings about there being exactly one nearest world), I would be able to divide the smallest measurement in half and get an even closer world. And then I could divide that smaller measurement, *ad infinitum*. Thus, Stalnaker's nearness relation would have the same problem as Zeno's Achilles.

Now that we have looked at the history of the discussion on counterfactuals we should move on to a more detailed analysis of Lewis' position. The above discussion on Lewis is quite terse, and a longer, more thorough explication will follow.

---

[21] Lewis (1981a), pp. 66-7.
[22] Ibid.
[23] Ibid., p. 60.
[24] Ibid., p. 63.

Chapter Two: Lewis, Possible World Semantics and Realism

In this chapter I will give a more thorough exposition of Lewis' modal realism. Then I will go on to discuss his justification for this system. I will then argue that his theory is not justified based on worries he himself raises, and responds to, in my opinion, unsatisfactorily. These worries have to do with the supposition which his whole theory rests on -- the existence of possible worlds.

1. Possible World Semantics

Lewis states that counterfactuals cannot be strict conditionals in the normal sense. Strict conditionals restrict access to possible worlds corresponding to a type of necessity. In modal logic the modal operators correspond to an accessibility relation. The accessibility relation indicates which worlds one can look to for truth conditions of the operators.[25]

Lewis introduces 'spheres of accessibility' to replace the accessibility relation. He states:

> It suits my purposes better not to use the customary accessibility relations, but instead to adopt a slightly different – but obviously equivalent – formulation. Corresponding to a necessity operator □, or a possibility operator ◊, or a kind of strict conditional, let us have an assignment to each world *i* of a set S_i of worlds, called a *sphere of accessibility* around *i* and regarded as the set of worlds accessible from *i*. The assignment of spheres to worlds may be called the *accessibility assignment* corresponding to the modal operators. It is used to give the truth conditions for modal sentences as follows.

[25] Lewis (1973), pp. 5-6.

> A sentence $\Box\phi$ is true at a world i if and only if ϕ is true throughout the sphere of accessibility S_i around i.
> A sentence $\Diamond\phi$ is true at a world i if and only if ϕ is true somewhere in the sphere S_i.
> A strict conditional $\Box(\phi \supset \psi)$ is true at i if and only if $\phi \supset \psi$ is true throughout the sphere S_i; that is, if and only if ψ is true at every ϕ-world in S_i.[26]

This interpretation is just Kripke's set of conditions.

To get to his ultimate purpose, giving truth conditions for counterfactuals, Lewis first discusses the continuum of strictness. Lewis points out that the more inclusive a sphere the stricter it is and the smaller it is the less strict it is. Thus, the logically strict conditional is the strictest and the vacuous conditional the weakest.[27]

Lewis argues that the counterfactual is a "kind of strict conditional based on comparative similarity."[28] He states that a counterfactual is true if the consequent obtains at all worlds where the antecedent obtains. He points out that not all worlds matter. Those worlds that are very different from ours would not give us much information about what could be here. He asserts that we would not look at worlds where Kangaroos used crutches to keep upright, to see if the counterfactual 'if Kangaroos had no tails, then they would topple over' is true. In a world where kangaroos knew how to use crutches and had access to them, many other things would be different as well.[29] We want to look to worlds where the least amount of change is necessary in order to make the counterfactual true.

A problem arises, however, if we interpret the counterfactual as an ordinary strict conditional. If we saw it as a strict conditional embedded counterfactuals, those with

---

[26]Ibid., p. 7.
[27] Ibid., p. 8.
[28] Ibid.
[29] Ibid., pp. 8-9.

more than one counterfactual claim, would eventually become vacuously true. Lewis illustrates this problem with a set of hypothetically embedded counterfactuals.

$\phi_1 \Box\!\!\rightarrow \psi$ and $\sim(\phi_1 \Box\!\!\rightarrow \sim\psi)$
$\phi_1 \& \phi_2 \Box\!\!\rightarrow \sim\psi$ and $\sim(\phi_1 \& \phi_2 \Box\!\!\rightarrow \psi)$
$\phi_1 \& \phi_2 \& \phi_3 \Box\!\!\rightarrow \psi$ and $\sim(\phi_1 \phi_2 \& \phi_3 \Box\!\!\rightarrow \sim\psi)$
.
.
.

> No one stage refutes the theory that the counterfactual is a strict conditional based on similarity, but any two adjacent stages do... If those at two adjacent stages both are true, then according to the theory the second is true vacuously. So are all those beyond it. Begin at the beginning: if ψ is true at every accessible ϕ_1-world but $\sim\psi$ is true at every accessible $(\phi_1 \& \phi_2)$-world, then there must not be any accessible $(\phi_1 \& \phi_2)$-worlds – nor any accessible $(\phi_1 \& \phi_2 \& \phi_3)$-worlds, nor ... Then if the lower counterfactuals are true, it is no thanks to their consequent.[30]

Thus, if we were to interpret the counterfactual as a strict conditional all embedded counterfactuals would be vacuously true. We use embedded counterfactuals often, e.g., 'If I win the lottery and if I have a huge tax bill, then I won't be rich.' This embedded counterfactual seems to be meaningfully true even though it has two counterfactual claims. The first counterfactual would be 'If I won the lottery, then I would be rich.' According to the theory we would look to sphere where I would be rich at all the worlds where I win the lottery. But, then the second counterfactual must kick in, adding to the claim that I would not be rich if I had a huge tax bill. According to Lewis there would be no accessible worlds in which I was both rich and had a huge tax bill if we were to interpret the conditional strictly because the sphere of accessibility would not change, thus the counterfactual would be vacuously true.

---

[30] Ibid., pp. 11-12. '$\Box\!\!\rightarrow$' is Lewis' symbol for the counterfactual conditional, officially read as "if it were the case that ___, then it would be the case that ...".

To deal with the problem arising from embedded counterfactuals Lewis maintains that the counterfactual does not have one strictness but many; he calls these conditionals variably strict conditionals.[31] He explains this variable strictness in terms of his system of spheres.

> Corresponding to a variably strict conditional... there must be an assignment to each world i of a set $\$_i$ of spheres of accessibility around i, some larger and smaller....
> Let \$ be an assignment to each possible world i of a set $\$_i$ of sets of possible worlds. Then \$ is a *(centered) system of spheres* and the members of each $\$_i$ are called spheres around i, if and only if, for each world i, the following conditions hold.
> (C) $\$_i$ is *centered in i*, that is, the set $\{i\}$ having i as its only member belongs to $\$_i$.
> (1) $\$_i$ is *nested*, that is, whenever S and T belong to $\$_i$, either S is included in T or T is included in S.
> (2) $\$_i$ is *closed under unions*; that is whenever S is a subset of $\$_i$ and $\cup S$ is the set of all worlds j such that j belongs to some member of S, $\cup S$ belongs to $\$_i$.
> (3) $\$_i$ is *closed under (nonempty) intersections*; that is, whenever S is a nonempty subset of $\$_i$ and $\cap S$ is a set of all worlds j such that j belongs to every member of S, $\cap S$ belongs to $\$_i$.[32]

The system of spheres is supposed to represent comparative similarity between worlds. The closer one gets to the world at which one is seeking truth the more similar to that world the worlds are. Lewis further explains why each of the criteria is necessary and justified. (C) he claims, is justified because every world is most similar to itself. One may think that an objection, along the lines of Lewis' objection to Stalnaker, might be appropriate here because if there are an infinite number of worlds there must be infinitely more similar worlds. But Lewis responds that the closest a world can be to another is exactly similar to the world in question, but even then, the world would be most like itself in the property of identity. (1) is needed because if some subset of spheres were not

[31] Ibid., p. 13.
[32] Ibid., pp. 13-14.

nested then there could be two worlds from different spheres within that subset which could both be closer, at different times, than the other, to the world in respect to their belonging to their respective spheres. Two worlds cannot both be closer than the other, thus, the spheres must be nested. (2) and (3) are needed for similar reasons as (1): worlds would be at the same time more similar to the real world, first for being within and out of the union then the intersection.

With all this apparatus in place Lewis proclaims that we are now ready for the truth conditions for counterfactuals.

> $\phi\Box\rightarrow\psi$ is true at a world i (according to a system of spheres \$) if and only if either
> (1) no ϕ-world belongs to any sphere S in $\$_i$, or
> (2) some sphere S in $\$_i$ does contain at least one ϕ-world, and $\phi\Box\rightarrow\psi$ holds at every world in S…
> In brief: a counterfactual is vacuously true if there is no antecedent permitting sphere, non-vacuously true if there is some antecedent-permitting sphere in which the consequent holds at every antecedent-world, and false otherwise.[33]

This is like the truth conditions for modal statements, with extra conditions behind the scenes, to make sure the problem cases are dealt with.

After Lewis sets up the truth conditions he wants to make the procedure simpler. To this end he introduces what he calls the Limit Assumption. This assumption is that given some antecedent A and some world there is a smallest antecedent-permitting sphere. Behind this assumption is the assumption that the nested spheres are nested in smaller and smaller spheres. Within the smallest sphere are the closest worlds.[34] The truth condition for counterfactuals, given the Limit Assumption, now becomes quite easy: "A counterfactual is true at i if and only if either (1) there is no antecedent-permitting sphere

[33] Ibid., p. 16.
[34] Ibid., pp. 19-20.

i, or (2) the consequent holds at every antecedent-world in the smallest antecedent permitting sphere around *i*."[35] Again, (1) is for the vacuous case and (2) for the normal case. But who is to say there is a closest sphere? If spheres are infinite there may be infinitely closer spheres. Lewis acknowledges this and states that without a smallest sphere, which would be the case given certain counterfactuals, the truth conditions "amount to this: if there are antecedent permitting spheres, then if we take smaller and smaller ones without end, eventually we come to one in which the consequent holds at every antecedent-world."[36] For example, if I was analyzing the counterfactual "If Elana were shorter than she is, then she would need to hem her pants" we would look to all the spheres where "Elana is shorter than she is." If we look to closer, smaller, spheres eventually there will be a sphere in which at all worlds where Elana is shorter than she is she also need to hems her pants.[37]

Lewis distinguishes between what he sees as two different types of counterfactuals: 'might' and 'would'.

> $\Box\!\!\rightarrow$
> read as *'If it were the case that ___, then it would be the case that ... '*,
> $\Diamond\!\!\rightarrow$
> read as *'If it were the case that ___, then it might be the case that ... '*.[38]

All of the prior truth-conditions were for the 'would' counterfactual. Given the 'would' counterfactual's truth conditions and given Lewis' definition of the might

---

[35] Ibid., p. 20.
[36] Ibid., p. 21.
[37] Given that the closer the sphere is to the truth seeking world the more similar the worlds in that sphere are to the target world it would be unlikely that Elana would need to hem her pants in the closest worlds. Thus, this counterfactual would probably be false if seen as a "would" counterfactual. "Would" and "might" counterfactuals are discussed in the next paragraph.
[38] Ibid., pp. 1-2.

counterfactual, $\phi \Diamond\!\!\rightarrow \psi =_{df.} \sim(\phi \Box\!\!\rightarrow \sim\psi)$, the truth conditions for the might counterfactual are:

> $\phi \Diamond\!\!\rightarrow \psi$ is true at a world *i* (according to a system of spheres \$) if and only if both
> (1) some ϕ-worlds belong to some sphere *S* in $\$_i$, and
> (2) every sphere *S* in $\$_i$ that contains at least one ϕ-world contains at least one world where $\phi \& \psi$ holds.[39]

Given the Limit Assumption the 'might' counterfactual would be true if the consequent obtained in some world within the closest antecedent-permitting sphere where the antecedent obtained.

2. Lewis' Defense

Lewis admits that his theory rests on suspicious assumptions. He thinks there are two things which a philosopher might object to: (1) the existence and or justification for asserting the existence of possible worlds, and (2) similarity between worlds. He states that these suspicions are not well founded.[40] Since only his defense of the existential properties of possible worlds is relevant to my later argument I will focus only on it.

Lewis starts his defense by asking "If our modal idioms are not quantifiers over possible worlds what else are they?"[41] He says that they cannot be unanalyzable primitives because this would not give an alternative to his theory but just be non-theorizing. He says they cannot be metalinguistic predicates analyzable in terms of consistency because "What is consistency?" Two sentences are consistent if, and only if, it is possible for both of them to be true. Thus, possibility cannot be analyzed in terms of

[39] Ibid., p. 21.
[40] Ibid., p. 84.
[41] Ibid., p. 85.

consistency or else it would lead to circularity. And, he states, they are not *so-called* possible worlds that are in fact linguistic entities.[42] They are, he argues, "respectable entities in their own right."[43]

Lewis raises some possible objections and then responds to them.

(1) Argument from Actuality: This objection is based on the fact that one wants to say that possible worlds do not *actually* exist. One wants to say there is this world and only this world exists. Lewis states that this is correct that he never implied that "unactualized worlds" *actually* exist. He claims that this does not imply that possible world realism is false. He states: "Realism about unactualized possibles is exactly the thesis that there are more things than actually exist."[44] Given what he takes the meaning of 'actuality' to be, he does not think this is problematic. An explication of this meaning will follow.

(2) Argument from Parsimony: Another worry might be that Lewis' theory implies the existence of too many entities. If one's theory is supposed to include as few entities as possible, then there is something wrong with Lewis' theory. Lewis responds that there are two types of parsimony: qualitative and quantitative. He adds that his theory is not qualitatively unparsimonious, but is quantitatively so. His reason for thinking that his theory is qualitatively parsimonious is because everyone believes in the existence of this world and he just asks us to believe in more entities like this one. He

---

[42] Lewis gives an extended argument against "ersatz realism" in Lewis (1986b). This takes place in a chapter called "Paradise on the Cheap?", pp. 136-91.
[43] Lewis (1973), p.86.
[44] Ibid.

states that this is no problem for him because he "recognizes no presumption whatever in favor of quantitative parsimony."[45]

(3) Argument from Mystery: Lewis mentions that there might be those that do not want to make any existence claims about possible worlds if they know very little about them. He admits that he would tell people more about possible worlds if they were creations of his, but since they are true entities he cannot. He states that he does not know much about them himself, nor does he know how to find out more. He knows they are entities like this world but does not know the specifics about them since they are causally independent of this world. He does however think that the theory is a good theory and thus is justified.[46]

Lewis thinks that a good theory ought to be both respectful of one's pre-philosophical beliefs as well as systematic. He explains:

> One comes to philosophy already endowed with a stock of opinions. It is not the business of philosophy either to undermine or to justify these preexisting opinions, to any great extent, but only to try to discover ways of expanding them into an orderly system… [A theory] succeeds to the extent that (1) it is systematic, and (2) it respects those of our pre-philosophical opinions to which we are firmly attached.[47]

Lewis then explains that he thinks his theory has the best fit with these criteria. He states that we have naïve beliefs that entities could have been other than they are, and that his theory is "the only successful attempt, that [he] know[s] of, to systematize these preexisting modal opinions."[48]

---

[45] Ibid., p. 87.
[46] Ibid.
[47] Ibid.
[48] Ibid.

3. Analysis

I have a problem with all of Lewis' defenses to the problems he raises. I do not think his defenses are sufficient.

Problems with Existence

Lewis claims that possible worlds really exist but not actually. Considering that Lewis wants to be faithful to our pre-philosophical opinions this seems strange. How does something, which does not actually exist, exist? The only way I could fathom is by reading Lewis as a Meinongian. Lewis might then mean that possible worlds exist like unicorns exist according to Meinong; they subsist. I believe the Meinongian concept of subsistence is wrong. Even if I thought it right there would still be a problem. If possible worlds subsist, à la Meinong, then how can Lewis just intuit that the rules governing them would be logical? I do not know if there are any rules governing subsistent objects and if there are I would not be so sure that they would be along the lines Lewis has in mind.[49]

Lewis does state, however, that he thinks 'actuality' is indexical. He states that we actually exist from our perspective and other-worldly-things do not. Otherworldly things actually exist from their own perspective, while from their perspective this-worldly things do not actually exist.[50] Indexicals are words like "I", "here" and "now." They are used to refer to different things/people in different contexts or depending on the time and person

[49]Meinong posited subsistent objects to serve as referents of objects that do not exist *per se* – like fictional characters or impossible objects. This was supposed to make sense of utterances which included references to things like unicorns or round squares. For more on whether or not Lewis is a Meinongian see Linsky and Zalta (1991).

[50] Lewis (1973), p. 86.

uttering them. Thus, Lewis thinks "actual" is indexical in the sense that my utterance of the sentence "this world is actual" means something different than if someone at another possible world were to utter it. This would be like the sentence -- "I am here" – "here" would refer to different places when uttered by people in different locations. This sounds a bit *ad hoc.* Why don't our pre-philosophical notions about actuality occur to him here? Pre-philosophically we believe that 'actuality' means just that: what really exists everywhere. He may think that 'actuality' is indexical because of questions about actuality over time; e.g., at the beginning of the twentieth century the statement "Bertrand Russell, the man who wrote about the man who wrote *Waverly*, actually exists", would be true, whereas now the statement would be false. Thus, actuality related to specific statements over time would be indexical. But what one can say about Russell's actual existence and possible worlds' actual existences is quite different. Russell *did* actually exist; he lived in this world.

Lewis might reply that my understanding of 'actual' is wrong. It might seem to me not indexical, but as I said myself with regard to time, we do use it indexically. My intuitions, and those like mine, then would just be wrong.

Parsimony

As stated above Lewis asserts that there are two types of parsimony: qualitative and quantitative. He states that while his theory is not parsimonious with regard to the quantitative parsimony it is qualitatively parsimonious. I agree that there are different types of parsimony, but I do not think that he passes the parsimony test.

First of all what is qualitative parsimony? Lewis does not explain what he means by qualitative parsimony. He does, however, give us a hint. Lewis states "My realism

about possible worlds is merely quantitatively, not qualitatively, unparsimonious. You believe in our world already. I ask you to believe in more things of that kind, not things of some new kind."[51] It seems he means by this that qualitative parsimony is keeping to a minimum new *types* of entities. Because we already have the world-type, and he is only adding more of the same type, he is, in fact, being parsimonious.

If we take a very narrow view of theory construction, maybe Lewis is right. We do know what a world is. But I think this is too narrow. We may have a world concept but with that concept comes the belief that our world is singular. The infamous stare, the use of "the" when talking about *the* world, the need for more explanation than just stating possible worlds are like this one, all of these facts are indications that our world concept includes uniqueness.

Lewis may not object. He may state that the concept of "world" does include uniqueness – *all worlds* ought to be unique. But this response would be missing the point. I am not merely talking about identity. I am talking about singularity. Our concept of the world includes the belief that our world is the only one like it. If this is true, the ability to add more world-like entities to our ontology would be impossible. How can we believe that there are many things that are of the same type which includes the condition that it be the only one of that type?

Mysteriousness

Lewis states that modal idioms exist, and they must be systematized, and his theory does so while being considerate of these idioms. First, Lewis seems to think that our modal idioms are translatable in terms of his possible worlds. This to me seems like

[51] Ibid., p. 87.

translating similes into possible world semantics and then stating that we are justified in saying that pretty-woman-x's face is like the moon because there is a world in which her face *is* the moon. Lewis would respond that we do not really believe that our use of metaphor indicates the existence of anything. We do not believe that those things, which we are comparing, are actually alike in *all* ways, just that they have *some* things in common.

There are many films in which the protagonist is thrown into a world/situation where things are other than they are. In *It's A Wonderful Life* Stewart's character George Bailey is shown a world in which he never existed.[52] Everything is different in Bedford Falls. No one in the town recognizes him, although before he was a well-known and respected man. George does not understand at first. He does not say "Ah, this is a possible world in which I never existed." Nor, does he say, "This is a way in which things could have been otherwise." Instead he accuses the angel of being first a drunkard then a hypnotist. This does not seem like a person who thought unactualized possibles do exist.

Lewis may respond that George did in fact believe that things could have been otherwise or else he would not have said "I wish I was never born." He initially did not believe the angel because one does not believe that one could have access to such worlds. George was not, then, disbelieving that the world without him existed but doubting whether he could view its existence.

Yet, Lewis would concede that the existence of possible worlds does not "do justice" to common sense. Lewis recognizes the "incredulous stare" phenomenon as an indication of this. He states that "[m]odal realism *does* disagree, to an extreme extent,

[52] I am indebted to Jason Holt who suggested this example to me.

with firm common sense opinion about what there is."[53] Thus, he does realize that possible worlds are strange and hard to accept. But, he falls back again on the virtue of systematization and how his theory's success in being systematic eases the worry of a common-sense backlash. Lewis acknowledges that the incredulous stare not only indicates that modal realism fails common-sense notions of ontology but that it indicates that others believe that it "fails the test" of weighing the cost of a theory, in this case dispensing with common sense notions of ontology, for the benefits. He disagrees. He adds:

> I acknowledge that my denial of common sense opinion is severe, and I think it is entirely right and proper to count that as a serious cost....*I* still think the price is right, high as it is. Modal realism ought to be accepted as true. The theoretical benefits are worth it. Provided that they cannot be had for less.[54]

Thus, it comes down to Lewis believing that his theory is the best theory systematically with the fewest costs. He makes it quite explicit that this does not mean that his theory is the only one, just that his is the best that there is.[55]

I have gone over the costs of Lewis' theory but what are the theoretical benefits, why is it the best systemically? Lewis answers this question in the first part of *On the Plurality of Worlds*. Lewis mentions four reasons why his theory is good. First, his theory is good because of its handling of modality. Modal realism's benefits with regard to systemic quantification over the necessary and possible are good.[56] Second, his theory handles counterfactual considerations well. Third, his theory is inclusive. With modal

---

[53] Lewis (1986), p. 133.
[54] Ibid., p. 135.
[55] Ibid., p. viii, (forward).
[56] Ibid., p. 5.

realism necessity restrictions are open to quantification with different types of necessity.[57] Fourth, one can quantify over properties.[58]

There is no set measure of how to weigh the benefits and costs of a theory but I have laid out what Lewis believes to be the costs and benefits of his theory. With that I will argue for a theory that may be comparable, if not better.

[57] Ibid., p. 27.
[58] Ibid., p. 50.

Chapter Three: Conditional Probability and Counterfactuals

Now that the preliminaries are done, I will be getting down to the real business. In this chapter I will be defending the claim that counterfactuals can be analyzed in terms of conditional probabilities. I will argue that probability, in this capacity, ought to have a frequentist interpretation. I will then argue that the conditional probability ought to be mapped on to a polyvalent system.

The thesis that conditionals can be analyzed in terms of conditional probability is not novel. Stalnaker defended a similar proposal with regard to indicative conditionals. Lewis argued that Stalnaker's thesis led to trivialities, and agreeing, Stalnaker abandoned it. I will discuss this debate and further discuss what I will be calling "The *Mind* Debate" which developed out of Lewis' triviality claim. I will tentatively argue, along with van Fraassen, that the triviality results only if one is a possible world realist, and since I am making no such ontological claim the triviality objection does not work against the conditional probability thesis. I will then argue that even if van Fraassen's objection fails counterfactuals are safe from the triviality claim. I will then finish off the chapter considering some possible problems.

Lewis asks "If our modal idioms are not quantifiers over possible worlds, then what else are they?"[59] Although Lewis seems very confident in his solution to the problem of counterfactuals there are other possible solutions. I will now discuss one such solution: conditional probability.

[59] Lewis (1976), p. 85.

1. Core Claim

As I stated I will defend the thesis that the truth-values for counterfactuals can be analyzed in terms of conditional probability. Where $p\Box\rightarrow q$ is a counterfactual:

(1) $p\Box\rightarrow q$ is true iff P(q/p) is sufficiently high.[60]

Thus, as a first blush, for any counterfactual to be true the probability that the consequent holds given that the antecedent holds must be sufficiently high.

This raises the obvious question: what do I mean by "sufficiently high"? I could state that anything above 0.5 (on a 0 to 1 probability scale) would be sufficiently high. But this would mean there would be many border line cases. On this analysis a counterfactual with the conditional probability of 0.501 would be sufficiently high yet 0.5 would not be. This would seem *ad hoc*. One could claim that anything below 0.5 is false and everything above 0.8 is sufficiently high and everything in between is indeterminate. This would be vague, but our use of conditionals is vague. There are those conditionals that one does not want to say are either true or false.

Given that the above description of counterfactual truth-conditions is too vague a more precise claim is necessary. Thus, given a multivalued logical system with a valuation function V whose range is the unit interval:

(2) $V(p\Box\rightarrow q) = P(q/p)$.

[60] Following standard notation, "P(q/p)" represents the *conditional probability* of q given p, i.e. the probability that q holds, given p.

This rendering of the claim entails that the conditional probability of the counterfactual mapped one-to-one onto a multivalued system. Thus, if the probability of the consequent given the antecedent were 0.6 the counterfactual would be 0.6 true.

One may be suspicious of my use of a polyvalent system. Yet, anyone who is willing to admit a third truth-value between true and false, indeterminate, will have to grant that there must be many values of indeterminacy. For, although I may not know whether or not such-and-such a proposition is true, and thus, say that it is indeterminate, I may have partial knowledge making the truth more or less likely. With counterfactuals I am claiming that when one would want to say that, in normal situations, something is indeterminate to some degree n, with counterfactuals that degree n represents the truth of that counterfactual.[61]

In the first rendition of the claim I said that the probability had to be sufficiently high for a counterfactual claim to be true. That claim would fit nicely if one was "stuck" with a bivalent system. If there were only two truth-values, true and false, then the probability of the consequent given the antecedent would have to be high enough for the counterfactual to be true. Although, the first gloss of my central claim may be the best one could do if one were stuck with the bivalent system, it is not sufficient because we can represent more about the truth with (2), including what would turn out to be borderline cases with (1).

A worry might be that if P(p) = 0, then P(q/p) would be undefined, i.e., if the antecedent of the counterfactual does not obtain the conditional probability of the counterfactual would not be defined. This would be an obvious worry for me since a lot

[61] This is assuming that the reason for invoking a multivalued system is for indeterminacy and not for other reasons, e.g., description of "more or less property p" with vague predicates.

of the time we utter counterfactuals when we know the antecedent is false. For example, I may say "If I were 6'2", then I would hit my head on the door frame." I am not 6'2", thus, the antecedent does not obtain. Yet, I would still think there ought to be a truth-value to my statement.

My solution to this involves the different interpretations of probability. Much of the current probability literature in philosophy presupposes a subjective account of probability. On this account probabilities represent subjects' degrees of belief in statements. This means that probability is dependent on individuals' thoughts and beliefs, whether they be true or not. Statements like "I feel statement x is true" would be common here. Instead of this interpretation, for the conditional probability of counterfactuals, I think that a frequentist interpretation of probability holds more promise. The frequentist interpretation takes all occurrences of event x and measures how often event y has occurred when x has. Although Ramsey defends a subjectivist interpretation of probability, he states that what we mean by "probability" in ordinary language most often does have a frequentist flavor.[62] What one means by "One has the probability 0.7 of getting sick to the stomach in Mexico if one drinks the tap water" is just that 70% of those who drink water out of the tap in Mexico get sick to the stomach. The percentage is then given in terms of probability, for the likelihood of those getting sick in the future under the same conditions given the past events. With counterfactuals I am appealing to occurrences in the past similar to the antecedent and consequent which will inform us of the truth-conditions for counterfactual statements.

Given the frequentist interpretation, a more precise rendering of the core claim

[62] Ramsey (1964), pp. 63-4.

can be made. 'If p were the case, then q would be the case' is roughly equivalent to the claim that the probability of q-situations obtaining given that p-situations obtain is quite high, i.e.:

(3) $p\Box\rightarrow q$ is true iff the probability of a q-situation obtaining given that a p-situation obtains is sufficiently high. (Given bivalence).
(4) $V(p\Box\rightarrow q)$ = the proportion of p-situations which are q-situations. (Given polyvalence).

So, if I say that "If I were 6'2", then I would hit my head on the door frame" I would mean that the percentage of those, in the past, who were 6'2" and who hit their heads on the doorframe in question (or doorframes of that measurement) was high. Thus, there would have to be certain restrictions, the person could not be bending over, the person cannot be in a wheel chair, the person cannot be brought over the threshold in someone's arms, etc., according to common sense.

There would be less clear situations where the frequency of q-situations obtaining given that a p-situation obtains would not necessarily be high even if everything else would be equal. Such cases would be akin to Lewis' "might" counterfactuals, where one might say "if you drink the tap water in Mexico, you might get sick." Here one would just use the equation in (2). To stay away from this ambiguity I think in analysis one ought to dispense with such subjective terms and keep with the "might" description.

Another problem that may be raised is that people may be unclear about what I mean by p- and q-situations. If I were to say, "If I were to scratch a cat's belly, she would purr", this would mean that my p-situations would be the set of all situations where cats were scratched on their belly and my q-situation would be the set of all occurrences of

cats purring. Then one sees how often cats have purred when they were scratched on their bellies.

This may be all fine and dandy but what about one-off situations? Say I break my arm, which has never occurred before, and I want to know how long it will take for my arm to heal. One may, at first blush think that my p-situation has a set of one since this would be the first time I have broken my arm. We must look at situations more widely. The sets ought to be comprised of situations that are similar. Thus, I might look to the set comprised of members who were all 28 year old females who drink very little milk and who have broken their arms. Or, I may look to the set of all people who have broken their arms and have similar breaks. Here we get into the problem of comparative similarities. How do we know which set to consult? When we are talking about similarities there are so many points of comparison.

Lewis also has the problem of comparative similarities when he sets up his system of nested spheres according to comparative similarities and I think his response to the problem is a good one. When discussing the problem he brings up the comparison of two cities:

> [A problem arises when] I say that Seattle resembles San Francisco more closely than it resembles Los Angeles. Does it? That depends on whether we attach more importance to the surrounding landscape, architecture [etc.]... Somehow we *do* have a familiar notion of comparative overall similarity.[63]

Lewis goes on to say that there are specific points of comparison that we use depending on what specifically we are comparing about the cities. If I am talking to you about the weather in different parts of the world and state that Moscow is more similar to Winnipeg

[63] Lewis (1976), p. 92.

than Kiev you would know that there are many points of comparison where Moscow and Kiev are more similar than Winnipeg and Moscow, the languages spoken in both Kiev and Moscow are Slavic, they were both under Soviet rule, etc. But you would ignore those similarities because you know that I am only comparing the weather and in that point Winnipeg and Moscow *are* more similar. For Lewis counterfactuals considered true in a certain context would depend on the speakers' interests in that context; with different comparative points the ordering of the spheres would change. With the broken arm case the set of p-situations would be determined by how accurate a probability one wants. If I wanted a more accurate probability I would look to the set of twenty-something-milk-deniers; if I wanted something more general I would look to the larger set. Thus, it becomes easier to determine which members to include in a set of p-situations when one has specific points of comparison in mind, when that is not sufficient one has desires about how accurate one wants the information, and thus, how wide or narrow the set ought to be. This may still be vague but as Lewis states in this context vagueness is not necessarily a bad thing because when one is dealing with a vague concept, like counterfactuality, maybe the apparatus used to determine the truth-conditions ought to be a little vague as well.[64]

I would like to make it clear that p-and q-situations are not *possibilia*. These "situations" are not possible situations but sets of actual occurrences. These are not "ways things could have been."

[64] Ibid., pp. 92-3.

2. The Stalnaker Thesis and Lewis' Triviality Claim

Stalnaker has forwarded a probability interpretation of conditionals. This theory came to be known as "The Stalnaker Hypothesis" or Stalnaker conditionals. He states that his reason for wanting to forward such an account is that although interpretations of probability are contentious its algebra is not, unlike the conditional.[65] His thesis states that the probabilities of c onditionals are just conditional probabilities, given a positive antecedent. Stalnaker thinks that the probability of "if x, then y" is just conditional probability; i.e., the probability of the consequent obtaining given the antecedent.

Lewis does not think that one can interpret conditionals in terms of conditional probability. He states that this "quest ... is futile."[66] Lewis argues that if one uses absolute probability absurdity results. He also shows that less than universal probability also results in absurdity. His first triviality proof is as follows:

Theorem: "Stalnaker's Hypothesis" [SH] together with the normal assumptions for the calculation of conditional probability results in triviality. Thus, along with the standard probability calculations:

(1) $P(C/A) =$ df $P(CA)/ P(A)$, if $P(A)$ is positive, i.e., $P(CA) = P(A) \times P(C/A)$. [the definition for conditional probability]

(2) SH: $P(A \rightarrow C) = P(C/A)$ if $P(A) > 0$.

(3) The Central Claim: $P(A \rightarrow C/B) = P(C/AB)$ if $P(AB) > 0$.

(4) $P(A \rightarrow C/C) = P(C/AC) = 1$. [From (3)]

(5) $P(A \rightarrow C/\sim C) = P(C/A\sim C) = 0$. [From (3)]

---

[65] Stalnaker (1981b), p. 107.
[66] Lewis (1981b), p. 130.

(6) P(D) = P(D/C) × P(C) + P(D/~C) × P(~C). [theorem of the probability calculus]

(7) P(A→C) = P(A→C/C) × P(C) + P(A→C/~C) × P(~C). [from 6]

(8) P(A/C) = 1 × P(C) + 0 × P(~C) = P(C).

(9) ∴ P(A→C) = P(C).[67]

What I have called the Central Claim [CC] is the step that the whole proof rests on. Once this is arrived at the rest of the proof follows. I will first explain the CC and then do a mini-proof for it.

The CC is arrived at by "conditionalization." Conditionalization means that "whenever P(B) is positive, there is a probability function P' such that P'(A) always equals P(A/B); *...P' comes from P by conditionalizing on B.*"[68]

Given the above explanation of conditionalizing a proof for (3) follows:

(1) P(A→C/B) = P'(A→C)

(2) = P'(C/A)

(2.1) = P'(AC)/ P'(A) [df. of conditional probability]

(2.2) =P(AC/B)/ P(A/B) [df. of P']

(2.3) =[P(ABC)/P(B)]/[P(AB)/P(B)] [df. of conditional probability]

(2.4) =P(ABC)/P(AB)

(3) = P(C/AB) [df. of conditional probability]

---

[67] Ibid., pp. 129-32. I have altered the way Lewis writes the proof. For the sake of brevity and clarity, I have omitted the general probability calculations, changed the numbers and some of the symbols.
[68] Ibid., p. 131.

The proof demonstrates that SH with CC leads to absurdity. This follows because the conditional probability would only be the probability of the consequent. One wants the probability of the conditional *given the antecedent*, thus, Lewis claims that equating the probability of a conditional with conditional probability does not work.

This proof may not be relevant to my thesis. Stalnaker's thesis and my core claim are different. It is claimed in Stalnaker's thesis that the probability of a conditional is equal to the probability of the consequent conditional on the antecedent, whereas it follows from my core claim that the degree of truth of the counterfactual is equal to the probability of the consequent on the antecedent. In other words, Stalnaker's thesis does not assert truth-conditions, whereas my core claim does. Lewis shows the SH entails $P(A \rightarrow C) = P(C)$. This would only be a problem for me if I held that the truth-value of a counterfactual is equal to its probability, which in is in turn fixed (as per SH) by the conditional probability; it would then follow via Lewis' proof, that the truth-value of a counterfactual is equal to the probability of its consequent, regardless of what the antecedent is. I can hold that the truth of the counterfactual is fixed by the conditional probability but at the same time deny that that probability maps one-to-one onto truth-values. For example, if a counterfactual has the conditional probability of 0.5 I can say that its truth-value is half true. This might avoid the problem but, as I stated before, I do want the probability to map one-to-one onto the truth-values. Not only does it cut out a step it is also quite intuitive. Therefore, the triviality proof does apply to my thesis and I ought to defend the thesis against it.

The triviality results would not faze some of those who think conditionals are disguised conditional probabilities, (e.g., Adams)[69], because they do not believe that indicative conditionals are governed by rules of truth but are governed by rules of assertability. Lewis responds that he does not have a conclusive argument against this position but his objection is that "the hypothesis requires too much of a fresh start."[70] By a fresh start Lewis presumably means that one ought not throw away fundamental pieces of logical assumptions (such as rules of truth) unless one has a very good reason for doing so. If the benefits outweigh the loss then it would be permitted (this is why Lewis believes *his* fresh start is justified).

Van Fraassen argues that the only reason triviality results for Lewis is because of his possible world realism. He states that Lewis is demanding "that the model structures be such that any probability assignment ... can be superimposed without violation of the [Stalnaker] Thesis, and without corresponding changes elsewhere in the model structure."[71] Essentially, Lewis believes that if a person receives new information or changes his mind the only thing that ought to be changed in the calculus is the probability measure and not the nearness of possible worlds. One would think that with new information the nearness of possible worlds would change too. But van Fraassen claims that Lewis' adherence to everything except the probability function results from his adherence to his metaphysics.[72] Van Fraassen states:

---

[69] Adams (1975).
[70] Lewis (1981b), p. 136.
[71] Van Fraassen (1976), pp. 273-74.
[72] Stalnaker (1976) calls this metaphysics "metaphysical realism" and defines it as follows: "The proposition expressed by a conditional sentence is independent of the probability function defined on it. So, for example, the content of the sentence B>C is the same in the context P(B>C) as it is in the context P_A(B>C)."

> The inspiration for the requirement must doubtlessly be Lewis' metaphysics, according to which one should always be able to say: let the possible worlds in my model structure be those which there actually are, let the nearness relation on them be the one reflecting their actual and objective similarities. In this scheme, the probability measure's nothing but a device to picture our ignorance. Hence, it has nothing to do with the internal constitution of the model structure, which is reality itself. For this reality of possible worlds exists independent of the mind, its evolution flows on its own tenor.[73]

Van Fraassen then goes on to dismiss the existence of possible worlds. Without possible worlds, van Fraassen claims, with new information our models and theoretical commitments change; only then would it be viable to change the probability function. He states that for those of us who believe in one world there is no problem of triviality for one does not have to claim that the only thing that changes when new information is obtained is the probability measure.

This might not be the clearest of objections but it may be promising. When we get new information not only do probability functions change, but so do other things. If we found out that the laws of physics were other than we believed them to be not only would this affect probability but it would affect other commitments. It might change our whole worldview. It can't be as simple as only changing probability functions.

With the triviality proof out of the way, whether or not van Fraassen's objection works, I can go on to other things without fear of triviality.

3. The *Mind* Debate

The above arguments were said to be about the indicative conditional, not subjunctive counterfactuals. Their difference is debated. Are they really different? What

[73] Lewis (1981b), pp. 274-5.

is that difference? First I would like to discuss what the difference is, then what this means to my thesis. A lot of the writing on this topic has been published in *Mind*; thus, I have dubbed this debate "The *Mind* Debate."

So what is this difference between indicative and subjunctive conditionals? Gibbard argues that there are two types of conditionals that vary in their semantics and their syntax. The first he calls "epistemic conditionals", these will correspond to indicative conditionals; the second he calls "nearness conditionals" these will correspond to counterfactuals. He states that the exemplar for the first is Adams and the second Stalnaker.[74] Gibbard presents two sentences:

(1) If Oswald hadn't shot Kennedy, no one would have shot Oswald.

(2) If Oswald didn't shoot Kennedy, no one shot Oswald.

The first is subjunctive the second indicative. Gibbard points out that in (1) "hadn't" is past perfect, whereas, the "didn't" in (2) is simple past. Both antecedents refer to a time prior to assertion. But this is just syntax, it does not tell us anything about the semantics.

Gibbard tells a story to illustrate the different meanings. There are three people on a ship: Sly Pete, Mr. Stone and Zack. Pete and Stone are playing poker. Stone has a good hand. Zack signals Pete about the contents of Stone's hand (time t^0). Stone being suspicious asks for the room to be cleared. Moments later (time t^1) Zack thinks about two conditionals:

(3) If Pete called, he won.

(4) If Pete had called, he would have won.[75]

---

[74] Gibbard (1981), pp. 222-3.
[75] Ibid., p. 227.

At t^1 Zack accepts (3), because he knows that Pete is a sly player who has the information about what is in Stone's hand. He knows Pete would only call if his hand was better than Stone's. Zack thinks (4) is probably false. Because he knows that Stone's hand is good he knows it is also unlikely that Pete would have a better hand. Thus, he thinks it probably false that if Pete had called he would have won. (3) is indicative, whereas, (4) is subjunctive.

Gibbard claims that the way to check indicatives is by a "Ramsey test."[76] If one wants to test "If A, then C" one hypothetically adds A to one's stock of beliefs and then sees if C follows. Gibbard states that "it is accepted by anyone whose corresponding conditional credence is sufficiently high."[77] This seems to be a rule of assertion. These rules govern when one can assert something rather than whether one is asserting truly. Thus, if Pete had called the game at t^0, and lost, Zack would not have been wrong in believing in (3) for his reasons for believing it were good.

Gibbard states that laws of truth rule the subjunctive conditional. He states that Stalnaker's treatment of counterfactuals (not Stalnaker's assumption/thesis) is the way to test these conditionals. He states the reason is that one's asserting has to do with the way the world is with the subjunctive conditional; whereas, with the indicative one is asserting based on what is in one's mind.[78] Thus, Gibbard is stating that indicatives are ruled by laws of assertion not truth contra subjunctive conditionals.

Gibbard considers another version of the Sly-Pete story. In this scenario another person is also present: Jack. At t^0 Jack sees both hands and sees that Pete has a low hand

[76] Based on the principles forwarded by Frank Ramsey in Ramsey (1964).
[77] Lewis (1981b), p. 227.
[78] Ibid.

while Stone has a high hand. At the same time Zack sees that Stone's hand is high and signals the content to Pete. He does not see Pete's hand. At t^1 Jack hands a note to someone that states: "If Pete called, he lost." At the same time Zack hands that person a note that states: "If Pete called, he won." The person who has these notes handed to her has an interest in knowing who won. She does not know who sent which note. Gibbard thinks that both were right. Both were assertible given Zack's and Jack's knowledge and beliefs. Zack knows that he signaled to Pete and Jack knows that Stone's hand is higher. Thus, Gibbard states, neither proposition is false, because neither one of the men have false beliefs.[79]

Based on Gibbard's thesis Edgington claims that counterfactuals do not have truth-values either.[80] She tells a story based on what she calls the "Gibbard Phenomenon", the phenomenon that occurs in the second Sly-Pete example.

There is a disease D, Vaccines A and B, and side-effect S. Neither A nor B completely prevent D. If you've had A and get D you get S; if you've had B and you get D you don't get S. If you've had both A and B you don't get D, thus, you don't get S. But it is very rare that people receive both A and B.

There are two observers X and Y and patient Jones. Both X and Y know that it is rare for people to get both A and B.

X knows that Jones has had A and thus is justified in believing that *if Jones gets D he will get S.*

Y knows that Jones has had B and thus is justified in believing that *if Jones gets D he will not get S.*

---

[79] Gibbard (1981), p. 231.
[80] Edgington (1995), pp. 293-6.

Each of their beliefs is justified by what information they have. They contradict each other, but finding out the whole truth won't show one is wrong since that includes the information that since Jones has had both vaccines he will not get D. The conclusion drawn is that the two conditionals cannot have truth-values.

Edgington then changes the story. Suppose that Jones was run over before he had the chance to get D, then X can say "If Jones had got D he would have got S" and Y can say "If Jones had got D he would not have got S." Thus, here the Gibbard phenomenon applies to counterfactuals. Thus, they too, Edgington concludes, do not have truth-values.

Morton in "Can Edgington Gibbard Counterfactuals?" states that Edgington is too quick with her belief that the counterfactuals in this example are similar to the indicatives in the Gibbard case. Morton interprets Edgington's argument as being one that states that when one has two indicatives that are Gibbard-y one can change them into their corresponding counterfactuals and the result would be Gibbard-y counterfactuals. He states this is mistaken. Morton claims that there are ways both counterfactuals can have opposite truth-values:

> Suppose that Jones almost missed the appointment to have vaccine A but vaccine B was administered at a time and place at which only extraordinary circumstances would have prevented Jones' presence. Then if Jones had got D it would have been because he did not get A, so that "if he had had D he would not have got S" is true, and "if he had had D he would have got S" is false.[81]

Morton goes on to say that other scenarios could be cashed out in which there are definite truth-values. He says, for example, if X and Y conferred on their information they would know that he had had both A and B and thus both their counterfactuals were false. I too

[81] Morton (1997), p. 102.

have problems with Edgington but my problems with Edgington stem from my problems with Gibbard.

First, I do not see the motivation for saying that indicatives are not ruled by truth. In Gibbard's example I would think that Zack was wrong in believing (3) if Pete had lost. Just because he is asserting something without full knowledge does not mean that one can't be wrong. Zack has certain information, based on that information he believes there is a certain likelihood of something resulting, he may be right about the likelihood but he is wrong about the fact. There is extra information here that may be important. Zack makes his claim about his belief in (3) based on his knowledge of Pete's skill. If Pete's skill is extremely high then if he loses given that he had information about Stone's hand there must have been something which happened which would also change the likelihood of (3) obtaining. Say that Pete, realizing that Stone was suspicious, after clearing the room, threw the hand in order to throw Stone off his sneaky path. Can Zack then claim that he was justified in believing (3)? Maybe he can *feel* justified, but he'd still be wrong. He either didn't take into account important information that he had or didn't know enough about sneakiness to make the claim. If he didn't take into account some important information then his calculation would have been wrong. If he didn't know enough about sneakiness to realize the resultant likelihood he would still be wrong. When one does not know to put in a certain important variable when doing a calculation one doesn't say he could not have been wrong because he didn't know to put in the variable, one says he is wrong for exactly that reason. Thus, I think the Gibbard case for using the Ramsey test for indicatives fails.

Gibbard's second Sly-Pete story I also believe to be problematic. As before there is a fact to the matter. Pete either won or lost. Not having enough information about the situation is not cause enough for saying that the two conditionals uttered do not have truth-values. The person who was handed the notes may not know which one to believe but that seems to be unimportant to the truth of the matter. I also do not think that Gibbard gave enough of the story. Both Jack and Zack have other beliefs or content to their conditionals which back-up (or not, as the case may be) their claims. Jack's added content included that they must not have changed their hands in any way for his conditional to be true. Zack, on the other hand must have believed that if Pete called he would have had high cards. One of these claims must be right. As before maybe Zack is not thinking about other possibilities. Maybe, again, Pete throws the hand.

If there is no fact of the matter, i.e., if Pete did not call, one may be inclined to say that neither Zack nor Jack is right or wrong but that both were still justified in their claims. I think that the defense against this would be much as the fact-of-the-matter scenario. Both have more in mind than just their conditional. It is irresponsible of them to phrase them as they did. Both should have said something like "If Pete called and..." Then their antecedents would not be identical and there would be no problem with them having opposite conclusions.

I think Edgington's Gibbard case also fails. Since it is a Gibbard case it will be much like above. Yet, I believe Y was correct. X should have known there was a certain likelihood that Jones had also had B. Given that all that X knew was that Jones had had A he should have made a different judgment. Maybe his judgment should have taken into account the small likelihood that Jones had had both. Even if the likelihood was minimal

it was not absent. This small chance should have entered into his belief, thus, making it doubly conditional. All that he could have said was "If Jones had only had vaccine A and got D then he would have had S." Y on the other hand was possibly silly in his phrasing. Since he knew that Jones had gotten B he knew that either Jones had also had A and would not get the disease, and thus would not get S, or if Jones had not also had A he might get the disease but he would not get S because of B. Thus, he should have suspended judgment about whether Jones would get D but stated that no matter what Jones would not get S. Because the likelihood was present that Jones did have both he should have taken this into account. Since I think there is no Gibbard phenomenon in Edgington's indicatives there would not be that problem with her counterfactuals.

There are three ways that I can go with regard to the counterfactual/indicative debate. (1) I could state, like Lewis, that indicatives are nothing more than material conditionals and as such have truth conditions (he also believes that they have assertability conditions which sometimes makes the truth conditions hard to see).[82] (2) I could follow Gibbard and claim that indicatives have assertability rules while counterfactuals have rules of truth. (3) I could state that there are no real differences between conditionals and indicatives and that both could be understood in terms of truth via conditional probability. I could then claim that the main difference between them in cases like the Oswald conditionals is due to changes in conditional probability over time.[83] I will suspend judgment for now. I have what I need. Counterfactuals in any case are truth-functional.

---

[82] Lewis first adopts Grice's account of assertability and then Jackson's. For his discussion on Jackson's view see Lewis (1986a).
[83] Edgington (1995), pp. 314-16.

4. Possible World Semantics vs. Probabilistic Semantics

Under scrutiny I think the probabilistic analysis fares better than possible worlds. The whole possible world ontology seems very unintuitive to me. How does one know what happens in these possible worlds? There exists no special trans-world telescope through which we could check our claims about the other worlds. It seems to me that this is all conjecture. Whereas, with probabilistic analysis one is not speaking about strange worlds that are inaccessible, one is speaking about probability which is already commonly held. Thus, the main problem with Lewis' modal realism is epistemological. With the frequentist interpretation one has epistemic access to the data, or at least one can have access. Lewis does not have this virtue. On Lewis' account since possible worlds are isolated from each other one does not have access to the truth. But one is supposed to know that they are right or wrong about their modal propositions given these worlds.

I also think the probability account does far less damage to one's common pre-philosophical ontological notions than Lewis'. To merely infer the existence of other worlds in which counterparts are walking around doing much what we do to explain common language use seems almost absurd. Lewis wants one to totally rearrange one's ontology. Probabilistic analysis, on the other hand, deals only with things common ontologies already accept, viz. probabilities. This is not a new concept. I am not positing the existence of strange un-seeable somethings for which I can only give vague explanations and absurd answers.

I think Lewis would state that his ontology of possible worlds actually does less damage to pre-philosophic notions than I might think. Lewis states, in fact, that we have pre-analysis conceptions of these worlds. Lewis states:

> It is uncontroversially true that things might be otherwise than they are. I believe, and so do you, that things could have been different in countless ways. But what does this mean? Ordinary language permits the paraphrase: there are many ways things could have been besides the ways they actually are . . . It says that there exist many entities of a certain description, to wit 'ways things could have been' . . . [T]aking the paraphrase at 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.[84]

Lewis then goes on to say that he believes these worlds (or 'ways things could have been') are real, as real as this world. Does he then mean that we all have a belief in possible worlds as he does?

I take offense at his presumptuousness about what others believe. If Lewis takes everything at face value in our ordinary language he may posit the existence of many strange entities. Does he believe that there exists a world in which the Pegasus exists? Probably. It is not uncontentiously true that things could have been otherwise. I do not think that it is so obviously controversial. I do not think that when one states something in ordinary language that their common-sense ontology must permit the existence of the things posited. In ordinary language one uses metaphor, but the metaphor does not imply that the thing is actually like the other thing. I do not believe that cream cheese is actually like a cloud, nor do I believe that I could have been otherwise in the sense that there is a possible me which exists and is otherwise.

---

[84]Ibid., p. 84.

In conclusion, I think I have proven that a conditional probability account (with a frequentist interpretation of probability) mapped onto a polyvalent logical system is a better account of counterfactuals than Lewis'. The probability account gives one access to data, is more in line with our pre-philosophical beliefs, and is not ontologically suspect. Lewis' account may have worked hypothetically, but if we are to feel any confidence in our ascription of truth to counterfactuals we must have a system that gives us more than hypothetical truth-values.

Bibliography

Adams, Ernest W., 1975: *The Logic of Conditionals*. Dordrecht: D. Reidel.

Chisholm, Rodrick M., 1946: "The Contrary-to-Fact Conditional". *Mind* 55, pp.289-307.

Edgington, Dorothy, 1995: "On Conditionals". *Mind* 104, pp. 235-330.

Freund, John E., 1973: *Introduction to Probability*. New York: Dover.

Gibbard, Alan, 1981: "Two Recent Theories of Conditionals" in Harper et. al. (eds.) 1981, pp. 211-47.

Goodman, Nelson, 1947: "Counterfactual Conditionals". *Journal of Philosophy* 44, pp. 113-128.

Harper, W.L. and Hooker, C.A. (eds.) 1976: *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science Vol. 1*. Dordrecht: D. Reidel.,

Harper, W.L., Stalnaker , R, and Pearce, C.T. (eds.) 1981: *Ifs*. Dordrecht: D. Reidel.

Kripke, Saul A., 1963: "Semantical Analysis of Modal Logic I Normal Modal Propositional Calculi". *Zeitschrift Für Mathematische Logik und Grundlagen der Mathematik* 9, pp. 67-96.

Lewis, David K., 1973: *Counterfactuals*. Cambridge: Harvard University Press.

Lewis, David K., 1981a: "Counterfactuals and Comparative Possibility" in Harper et. al. (eds.) 1981, pp. 57-86.

Lewis, David K., 1981b: "Probabilities of Conditionals and Conditional Probabilities" in Harper et. al. (eds.) 1981, pp.129-150.

Lewis, David K., 1986a: "*Postscript* to 'Probabilities of Conditionals and Conditional Probabilities'", *Philosophical Papers Volume II*. Oxford University Press. Pp. 152-6.

Lewis, David K., 1986b: *On the Plurality of Worlds*. Oxford: Blackwell.

Lewis, David K., 1991: "Probabilities of Conditionals and Conditional Probabilities II". *Philosophical Review* 5, pp. 581-9.

Linsky, Bernard and Edward Zalta, 1991:"Is Lewis a Meinongian?". *Australasian Journal of Philosophy* 69, pp.438-53.

Morton, Adam, 1997: "Can Edgington Gibbard Counterfactuals?". *Mind*, pp.101-5.

Pollock, John, 1976: *Subjunctive Reasoning.* Dordrecht: D. Reidel.

Quine, Willard Van Orman, 1960: *Word and Object.* Cambridge: MIT Press.

Ramsey, Frank Plumpton, 1964: "Truth and Probability" in *Studies in Subjective Probability.* New York: John Wiley, pp. 61-92.

Stalnaker, Robert C., 1976: "Letter by Stalnaker to van Fraassen: January 22 1974" in Harper and Hooker (eds.) 1976, pp. 302-6.

Stalnaker, Robert C., 1981a: "A Theory of Conditionals" in Harper et. al. (eds.) 1981, pp. 41-56.

Stalnaker, Robert C., 1981b: "Probability and Conditionals" in Harper et. al. (eds.) 1981, pp. 107-128.

Van Fraassen, Bas C., 1976: "Probabilities of Conditionals" in Harper and Hooker (eds.) 1976, pp. 261-308.

Warmbrod, Ken, 1999: *Logic and Philosophy of Logic.* Winnipeg: U Manitoba.