Conditional and Counterfactual Logic

William B. Starr

To appear in Knauff, M. & Spohn, W. (in press). *The Handbook of Rationality*. Cambridge, MA: MIT Press

Abstract

A logic aims to capture certain rational constraints on reasoning in a precise formal language that has a precise semantics. For conditionals, this aim has produced a vast and mathematically sophisticated literature that could easily be the topic of a whole volume of this size. In this chapter, the focus will be on covering the major logical analyses developed by philosophers, the key issues that motivate them, and their connection to views developed in artificial intelligence, linguistics and psychology.

1 Introduction

Research on conditionals makes use of a few key terms and notation:

Conditional A sentence of the form *If A then B*, and its variants.

Antecedent The *A* component of a conditional

Consequent The *B* component of a conditional

Notation *If A then B* is represented in logics as $A \to B$, where A and B are logical representations of two sentences A and B

Conditionals are typically divided into two broad classes:[1]

Indicative Conditionals E.g. *If Maya sang, then Nelson danced*

Counterfactual Conditionals E.g. *If Maya had sung, then Nelson would have danced*

A crucial dividing line between indicatives and counterfactuals is that counterfactuals can be used felicitously to talk about situations where the antecedent is contrary-to-fact (or thought to be false) (Stalnaker 1975; Veltman 1986). This contrast is evident in (1) and (2).[2]

(1) Maya has definitely never sang. #If Maya sang, then Nelson danced.

(2) Maya has definitely never sang. If Maya had sang, then Nelson would have danced.

[1]While commonly assumed, this division is debated (e.g. Dudman 1988). Others prefer the categories of indicative and *subjunctive* conditionals (e.g. Declerck and Reed 2001: 99, von Fintel 1999). For an overview of these issues see Starr (2019). For a dedicated survey of subjunctive conditionals see von Fintel (2012). For indicatives see Gillies (2012) and Edgington (2014).

[2]Here, the '#' is used in the descriptive conventions of linguistics, where it indicates the native speaker judgment that the sentences are grammatical, but can't be used in this context.

Corresponding indicative and counterfactual conditions can also differ in truth-value Lewis (1973: 3), or at least in their justification (Adams 1975). Assuming Oswald was a lone shooter, the indicative (3) is straightforwardly true and justified, while the counterfactual (4) is false, or at least unjustified.

(3) If Oswald didn't kill Kennedy, someone else did.

(4) If Oswald hadn't killed Kennedy, someone else would've.

It is worth emphasizing that the term *counterfactual* should not be taken too literally, as it does not mean that a sentence of this form *must* have a contrary-to-fact antecedent (Anderson 1951). For example, the counterfactual (5) is used as part of an argument that the antecedent is true:

(5) If there had been intensive agriculture in the Pre-Columbian Americas, the natural environment would have been impacted in specific ways. That is exactly what we find in many watersheds.

A logic of conditionals typically aims to say which arguments involving $A \to B$ are deductively valid.[3]

Deductive Validity An argument with premises $P_1, \ldots, P_n$ and conclusion C is *deductively valid* just in case it is impossible for the premises to be true while the conclusion is false. *Notation*: $P_1, \ldots, P_n \vDash C$.

However, as discussed in §4, some instead follow Adams (1975) and focus on inductive support: whether the premises being true, or highly probable, makes the conclusion *highly probable*. Either of these approaches counts as pursuing a *semantic* approach to the logic of conditionals. The semantic approach well-suited to capturing the pervasive *context sensitivity* of conditional reasoning, which will be a prominent theme of this chapter.

Section 2 will outline the major challenges for a logic of conditionals and explain why classical truth-functional logic is not adequate. This will motivate §3 which discusses three different analyses that draw on tools from modal logic: strict conditional analyses, similarity analyses and restrictor analyses. This section will also outline how these analyses have been developed using a different approach to semantics: dynamic semantics. In section §4, the chapter will turn to analyses that rely instead on the tools of probability theory.[4]

2 Logic, Conditionals and Context

How can we systematically specify what the world must be like if a given conditional is true and thereby capture patterns of valid deductive arguments involving them? It turns out to be rather difficult to answer this question using the tools of classical logic. Seeing why will help identify the key challenges for a logic of conditionals.

The logical semantics developed by Frege, Tarski and Carnap provided useful analyses of English connectives like *and*, *or* and *not* using the Boolean truth-functional connectives $\wedge$, $\vee$ and $\neg$. In truth-functional semantics (see Chapter 3.1) the meaning of a sentence is identified with its truth-value True (1) or False (0), and the meaning of a connective is identified with a function from one

[3]This chapter will assume that the *if... then...* structure in indicative and counterfactual conditionals should have the same logical analysis and that their difference should be explained in terms of their different morphology. Not all philosophers have shared this assumption (e.g. Lewis 1973), but it remains the default one for good reasons. See Starr (2019) for further discussion.

[4]For a more exhaustive and formal survey of conditional logics see Arlo-Costa and Egré (2016).

or more truth-values to another, as depicted in Table 1. The best truth-functional approximation of

A	B	$\neg A$	$A \wedge B$	$A \vee B$	$A \supset B$
1	1	0	1	1	1
1	0	0	0	1	0
0	1	1	0	1	1
0	0	1	0	0	1

Table 1: Negation ($\neg$), Conjunction ($\wedge$), Material Conditional ($\supset$) as functions of True (1)/False (0)

if... then... is the material conditional $\supset$. $A \supset B$ is false when A is true and B false, and it is true otherwise (making it equivalent to $\neg A \vee B$). This analysis is employed in most introductory logic textbooks because it captures three key logical features of conditionals:[5]

Modus Ponens $A \rightarrow B, A \vDash B$
If a conditional and its antecedent are true, its consequent must be true.

No Affirming the Antecedent $A \rightarrow B \nvDash A$
A conditional can be true even when its antecedent is not.

No Affirming the Consequent $A \rightarrow B \nvDash B$
A conditional can be true even when its consequent is not.

The intuitive appeal of modus ponens is clear in (6-a). (6-b) and (6-c) can be indirectly confirmed by showing that a conditional can be consistent with denying its antecedent or consequent.

(6) a. If Maya sang, then Nelson danced. Maya did sing. Therefore, Nelson danced.
b. If Maya sang, then Nelson danced. But Maya did not sing.
c. If Maya sang, then Nelson danced. But Nelson didn't dance.

To see that modus ponens is valid for $A \supset B$ note that when $A \supset B$ is true and A is true, B must be true — as rows 1, 3 and 4 of Table 1 show. Row 4 shows that it is possible for $A \supset B$ to be true without either A or B being true.

Despite having some attractive features, the material conditional analysis is widely taken to be incorrect.[6] There are particular problems for analyzing English conditionals as $\supset$. But there are also general problems with *any* truth-functional analysis. Let us consider the more general problems first.

Many counterfactuals have false antecedents and consequents, but some are true and some false. (7-a) is false — given that Joplin's "Mercedes Benz" was a critique of consumerism — and (7-b) is true.

(7) a. If Janis Joplin were alive today, she would drive a Mercedes-Benz.
b. If Janis Joplin were alive today, she would metabolize food.

This is not possible on a truth-functional analysis: the truth-values of antecedent and consequent determine a unique truth-value for the whole sentence. With a bit of care, a similar point can be made about indicatives.

Suppose a standard die has been tossed, but you do not know which side has landed face up. (8-a) is intuitively true, while (8-b) is false.

[5] As is often remarked in those textbooks, it is the *only* truth-function that captures these features.
[6] While Grice (1989) attempted a pragmatic defense, it has never been satisfactorily extended to the problems presented here.

(8) a. If the die came up 2, it came up even.
b. If the die came up 1, it came up even.

This intuition persists even when you get to see that the die came up 3. It would seem that (8-a) is true and (8-b) false even though both have a false antecedent and consequent.

Another kind of problem for truth-functional analyses centers on the context-sensitivity of conditionals. The basic observation is that the truth-value of a conditional can vary from one context of use to another, even when the truth-values of the antecedent and consequent stay the same across the two contexts. This is clearest with counterfactuals.

Quine (1982: Ch.3) voiced skepticism that any semantic analysis of counterfactuals was possible by highlighting puzzling pairs like (9).

(9) a. If Caesar had been in charge [in Korea], he would have used the atom bomb.
b. If Caesar had been in charge [in Korea], he would have used catapults.

But Lewis (1973: 67) took these examples to show that the truth-conditions of counterfactuals are **context-sensitive**. The antecedents and consequents of (9-a) and (9-b) are all false, but in some conversational contexts (9-a) seems true and (9-b) false, and in other conversational contexts (9-a) seems false and (9-b) true. Consider evaluating (9-a) and (9-b) in a context where we have explicitly discussed and agreed that Caesar was, first and foremost, a ruthless military leader. In such a context, (9-a) seems like a true thing to say, while (9-b) seems false. By contrast, consider a context where we have explicitly discussed and agreed that Caesar was, first and foremost, a technologically conservative military leader. Then (9-b) seems like a true thing to say, while (9-a) seems false.

Similar examples illustrate the same point for indicative conditionals. Suppose we have mutually established that a die in our possession has 3 on every side except for one, which has 2. The die has been tossed, but we do not know how it came up. (10) seems true.

(10) If the die came up even, it came up 2.

Suppose we then find out that the die came up 3. It seems that (10) was still a true thing to say. However, consider a more ordinary context where we have mutually established that our die is a standard one. The die has been tossed, but we do not know how it came up. It would be false to say (10). Now suppose we later find out that the die came up 3. It still seems like (10) is false. This shows that the truth-value of (10) can vary from context to context, even when the truth-values of its antecedent and consequent are held fixed.

These general problems for a truth-functional analysis are compounded by particular weaknesses in the material conditional analysis. This is particularly clear for counterfactuals. As rows 3 and 4 of Table 1 makes clear, $A \supset B$ is true any time A is false. This mean that the material conditional validates this logical principle.

Material Antecedent $\neg A \vDash A \supset B$

But this means that *all* truly contrary-to-fact conditionals are true on a material conditional analysis. So not only is (11-a) in correctly predicted to be true, but *both* (11-a) and (11-b) are predicted to be true despite the fact that they seem to be contradictory.

(11) a. If the Earth hadn't existed, the moon would have existed.
b. If the Earth hadn't existed, the moon wouldn't have existed.

For indicatives this problem surfaces in another way: recall that indicatives are not felicitous to use when their antecedent has been explicitly denied. This is quite puzzling if that pattern of use is actually a valid form of argument.

Another major problem for the material conditional analysis stems from the fact that it validates:

Material Negation $\neg(A \supset B) \vDash A$

Neither of the following are compelling arguments for the existence of God, despite having (plausibly) true premises:

(12) a. It's not true that if God exists, he's a turnip. Therefore, God exists.
b. It's not true that if God had existed, he would be a turnip. Therefore, God exists.

But the shortcomings of this analysis are instructive, as they establish some clear criteria for a more successful analysis:

Non-Truth-Functionality The truth of a conditional is not determined by the truth of its antecedent and consequent.

Context Dependence The truth of a conditional depends on certain features of the context in which it is used.

Logical Constraints Conditionals do not obey Material Antecedent or Material Negation.

This is the starting point for analyses that appeal to *possible worlds*.

3 Conditionals and Possible Worlds

Table 1 helps make salient a crucial assumption of truth-functional semantics: the truth-value of a complex sentence is determined only by the truth-values of its component sentences *in that row*. Once the truth of A and B have been settled, the truth of $A \wedge B$ has been settled. But Table 1 also makes salient an alternative approach: what if the truth-value of a complex sentence is determined by the *distribution* of their truth-values across a number of rows?

Pursuing this alternative analysis requires clarifying exactly what 'a row' is, and settling on a particular account of which distributions matter. Following the tradition in modal logic, one can think of the rows as *possible worlds* (Kripke 1963). The two basic analyses surveyed in this section — strict-conditional analyses (§3.1) and similarity analyses (§3.2) — amount to different views about which distributions of truth-values across worlds make conditionals true. Sections 3.3 and 3.4 discuss two resources for extending these two kinds of analyses. Before exploring the analyses in detail, it is useful to state their shared assumptions and their main difference.

Strict and similarity analyses both employ the idea of a *possible worlds* w, which is simply a way the world could be, or could have been. They are treated as primitive points in the set of all possible worlds W. They play a crucial role in assigning truth-values to sentences: a sentence A is only true given a possible world w, but since w is genuinely possible, it cannot be the case that both A and $\neg A$ are true at w. Consider, then, a language with just three atomic sentences A, B, C, i.e. *Maya ate apples*, *Maya ate bananas* and *Maya ate cherries*. At least 8 possible worlds are then needed, as listed in Table 2:[7] Strict and similarity analyses both aim to say when a conditional like $A \rightarrow B$ is true in a given world w, and do so on the basis of whether certain A-worlds *relevant* to w are also B-worlds. But they differ terms of how they determine this set of worlds relevant to w.

Different methods for selecting relevant worlds produce a distinctive logical difference between strict and similarity analyses. This logical difference is easiest to illustrate with this principle:

[7] For technical reasons not discussed here, it is generally assumed in modal logic that there are more possible worlds than unique combinations of atomic truth-values.

	A	B	C
w_1	1	1	1
w_2	1	1	0
w_3	1	0	1
w_4	1	0	0
w_5	0	1	1
w_6	0	1	0
w_7	0	0	1
w_8	0	0	0

Table 2: Possible Worlds for A, B, C

Antecedent Strengthening $A \to C \vDash (A \wedge B) \to C$

Goodman (1947) observes that this does not hold for counterfactuals, since (13-a) is true and (13-b) false.

(13) a. If I had struck this match, it would have lit.
$S > L$
b. If I had struck this match and done so in a room without oxygen, it would have lit.
$(S \wedge \neg O) > L$

Lewis (1973: 10) dramatized this counterexample by considering 'Sobel sequences' such as (14). Surprisingly, adding more information to the antecedent repeatedly flips the truth-value of the counterfactual.

(14) a. If I had shirked my duty, no harm would have ensued.
$I > \neg H$
b. Though, if I had shirked my duty and you had too, harm would have ensued.
$(I \wedge U) > H$
c. Yet, if I had shirked my duty, you had shirked your duty and a third person done more than their duty, then no harm would have ensued.
$(I \wedge U \wedge T) > \neg H$
$\vdots$

For indicatives, (15-a) may be true when (15-b) is not.

(15) a. If Maya sang at the party, then Nelson danced at the party.
b. If Maya sang at the party and Nelson wasn't there, then Nelson danced at the party.

Similarity analyses predict that antecedent strengthening is invalid while strict analyses predict that it is valid. Strict analyses address the counterexamples above *pragmatically*: the conclusions of the arguments are false, but that is because pragmatic mechanisms force the conclusion to be interpreted relative to a different set of worlds than the premises. In doing so, these analyses appeal to a feature already highlighted above: conditionals' context-sensitivity.

The general pattern, of which antecedent strengthening is an instance, is:

Antecedent Monotonicity If $B \vDash A$ then $A \to C \vDash B \to C$

Another instance of this is:

Simplification of Disjunctive Antecedents (SDA)
$(A \vee B) \to C \vDash (A \to C) \wedge (B \to C)$

Antecedent monotonicity also leads (indirectly) to (Starr 2019: §2.1):

Transitivity $A \to B, B \to C \vDash A \to C$

Contraposition $A \to B \vDash \neg B \to \neg A$

Counterexamples similar to those above have been presented by similarity theorists to transitivity, contraposition and SDA (see e.g. Stalnaker 1968, Lewis 1973, McKay and van Inwagen 1977). Strict theorists have also aimed to address these examples pragmatically (e.g. Warmbrōd 1981). So the key issue dividing strict and similarity issues is this:

Pragmatic or Semantic Non-Monotonicity? Is the non-montonicity of conditional antecedents best explained semantically or pragmatically?

- *Strict Theorists*: pragmatically (e.g. Warmbrōd 1981; Gillies 2007, 2009)
- *Similarity Theorists*: semantically (e.g. Lewis 1973; Stalnaker 1968)

This chapter will not weigh in on this question, instead presenting the basics of these two approaches. Antecedent strengthening will be the focus for concreteness, but it is just a representative of antecedent monotonicity more generally.

3.1 Strict-Conditional Analyses

Strict conditional analyses began with the basic idea that $A \to B$ is true just in case all of the A-worlds are B-worlds (Peirce 1896). This has been refined using a crucial tool from the semantics of modal logic: an *accessibility function* $R(w)$. $R(w)$ takes a world w and returns the set of worlds that are accessible from, or relevant to, w (Kripke 1963).[8]

Strict Conditional
$A \strictif B$ is true in w, given R, just in case every A-world in $R(w)$ is a B-world.

To illustrate this definition, consider a particular world w_7 and a particular accessibility function R_1 where $R_1(w_7)$ corresponds to the set of worlds bolded in Table 3. $A \strictif C$ comes out true in this case: every bold world where A is 1, is a world where C is 1. $A \strictif C$ would come out false on a slightly different accessibility function R_2, which is just like R_1 except it includes w_2: a world where A is 1 and C is 0. R_2 in fact shows why antecedent negation does not hold for the strict conditional: $\neg A$ is true in w_7 and yet $A \strictif C$ is false in w_7, relative to R_2. Similarly, this is also a counterexample to material negation since $\neg(A \strictif C)$ is true in w_7, relative to R_2, and A is false in w_7. These count as counterexamples given the following definition of validity.

Modal Validity $P_1, \ldots, P_n \vDash C$ just in case for every w and R, if $P_1, \ldots, P_n$ are true in w, relative to R, then C is true in w, relative to R.

[8]Kripke (1963), and most work in modal logic, actually uses accessibility *relations* $R(w, w')$. But accessibility functions simplify the presentation, and are definable in terms of accessibility relations: $R(w) = \{w' \mid R(w, w')\}$.

	A	B	C
$\boldsymbol{w}_1$	**1**	**1**	**1**
w_2	1	1	0
$\boldsymbol{w}_3$	**1**	**0**	**1**
w_4	1	0	0
w_5	0	1	1
w_6	0	1	0
$\boldsymbol{w}_7$	**0**	**0**	**1**
$\boldsymbol{w}_8$	**0**	**0**	**0**

Table 3: Possible Worlds for A, B, C, Worlds in $R_1(w_7)$ in **bold**

According to this definition, antecedent strengthening is valid, an issue to which we will return. Modus ponens is valid for the strict conditional if one requires that accessibility functions satisfy *reflexivity*: for all w, w is in $R(w)$. Accordingly, the strict conditional addresses the main *logical* problems faced by the material conditional analysis. As discussed at the outset of this section, the strict analysis also captures non-truth-functionality: the truth-value of $A \mathrel{-\!\!\!3} B$ depends not just on the actual truth-values of A and B, but on their truth-values across the worlds in $R(w)$.

The context-sensitivity of conditionals is central to contemporary strict analyses, particularly when responding to examples like (13), (14) and (15). These examples seem to show that antecedent strengthening is not valid for conditionals, contrary to the strict analysis. While early work on the logic of strict conditionals did not include this component, subsequent work like Warmbrōd (1981) has done so by treating $R(w)$ as one aspect of context that changes as utterances take place.

Warmbrōd (1981) proposes that the accessibility function $R(w)$ corresponds to the background suppositions of the conversationalists, and will change as the conversation unfolds. In particular, Warmbrōd (1981) proposes that $R(w)$ also changes as the result of conditionals being asserted. Warmbrōd (1981) notes that *trivial* strict conditionals are not pragmatically useful in conversation. A strict conditional $A \mathrel{-\!\!\!3} C$ is trivial just in case A is inconsistent with $R(w)$ — there is no A-world in $R(w)$. Asserting $A \mathrel{-\!\!\!3} C$ in such a context does not provide any information: every conditional of the form $A \mathrel{-\!\!\!3} X$ is true. But Warmbrōd (1981) proposes that conversationalists adapt a pragmatic rule of charitable interpretation to make sense of why the speaker asserted $A \mathrel{-\!\!\!3} C$ rather than a different conditional:

(P) If the antecedent A of a conditional is itself consistent, then there must be at least one A-world in $R(w)$.

Given this, $R(w)$ can change as a result of asserting a conditional. This part of the view is central to explaining away counterexamples to antecedent monotonic validities, like antecedent strengthening.

Consider again this counterexample to antecedent strengthening.

(13) a. If I had struck this match, it would have lit.
b. If I had struck this match and done so in a room without oxygen, it would have lit.

On a strict analysis, if (13-a) is true in a world w, relative to R_0, then $R_0(w)$ must exclude any worlds where the match is struck but there is no oxygen in the room. However, if (13-b) is interpreted against R_0, the antecedent will be inconsistent with $R_0(w)$ and so express a trivial strict conditional. According to Warmbrōd (1981), interpreting (13-b) according to (P) forces the conversationalist to adopt a new, modified accessibility function $R_1(w)$ where the presence of oxygen is no longer assumed. If this is right, then (13) is not really a counterexample to the validity of antecedent strengthening: the premise is true relative to R_0 and the conclusion is only false relative to R_1. Warmbrōd (1981), and others, develop versions of this account to address all of the proposed counterexamples to antecedent monotonic patterns.[9] These accounts do have an important limitation: they do not capture nested conditionals, and do not actually predict how $R(w)$ evolves to satisfy (P). von Fintel (2001), Gillies (2007, 2009) and Willer (2017) offer accounts that remove these limitations, using the tools of dynamic semantics covered in §3.4.

3.2 Similarity Analyses

The basic idea of similarity analyses is that $A \to C$ is true in w when C is true in all of the A-worlds most similar to w (Lewis 1973; Stalnaker 1968; Stalnaker and Thomason 1970). One way of precisely formulating this appeals to a selection function f, which takes a world w, a proposition p and returns the set of p-worlds most similar to w: $f(w,p)$.[10] This is used to define the truth-conditions of a similarity-based conditional, notated '>', as follows:[11]

Similarity Analysis

$A > C$ is true in w, relative to f, just in case every world in $f(w,A)$ is a C-world.

While a strict analysis assumes a *single set of relevant worlds* $R(w)$ for all conditionals, f can select a *different set of worlds for each different antecedent*. For example, it is perfectly possible for $f(w, A \wedge B)$ to contain worlds which are not in $f(w,A)$, even though all $A \wedge B$-worlds are A-worlds. Different similarity analyses propose different constraints on f. The candidate constraints are given in Table 4.[12] To make clear that this permits $f(w, A \wedge B)$ to contain worlds not in $f(w,A)$, and how

(a)	$f(w,p) \subseteq p$	**success**
(b)	$f(w,p) = \{w\}$, if $w \in p$	**strong centering**
(c)	$f(w,p) \subseteq q$ & $f(w,q) \subseteq p \implies f(w,p) = f(w,q)$	**uniformity**
(d)	$f(w,p)$ contains *at most* one world	**uniqueness**

Table 4: Candidate Constraints on Selection Functions; $p, q \subseteq W$ and $w \in W$

this invalidates antecedent strengthening, let's consider a concrete example.

Consider the worlds and selection function f_1 in Table 5. $A > C$ is true in w_6, relative to f_1, since A and C are true in w_3. But $(A \wedge B) > C$ is false in w_6, relative to f_1, since $A \wedge B$ is true and C false in w_2. It is easily verified that f_1 satisfies all four constraints in Table 4. This should highlight the key difference between strict and similarity analyses: strict analyses assume a fixed set of relevant worlds for all antecedents, while similarity analyses allow the set of relevant worlds vary from antecedent to antecedent — even among logically related antecedents. It is also worth highlighting a point that will matter later: nothing in the formal analysis requires $f(w,p)$ to hold fixed facts of w, even if they are unrelated to p. This is clear with $f_1(A, w_6) = \{w_3\}$ which does not preserve the fact that B is true in w_6 and B's truth may be unrelated to A being false in w_6.

[9]They also address the 'paradoxes of strict implication', e.g. $\neg \Diamond A \vDash A \dashv C$.

[10]See Lewis (1973: §2.7) for the various formulations. The set selection formulation makes the limit assumption: A-worlds do not get indefinitely more and more similar to w. Lewis (1973) rejected this assumption, but it will merely simplify exposition here. See Starr (2019) for discussion of the limit assumption.

[11]$f(w,A)$ is shorthand for $f(w, [\![A]\!])$, where $[\![A]\!]$ is the set of worlds in which A is true.

[12]Pollock (1976) adopts (a) and (b), Lewis (1973) and Nute (1975) adopt (a)–(c), and Stalnaker (1968) adopts (a)–(d).

	A	B	C
w_1	1	1	1
$\underline{w_2}$	$\underline{1}$	$\underline{1}$	$\underline{0}$
$\boldsymbol{w_3}$	**1**	**0**	**1**
w_4	1	0	0
w_5	0	1	1
w_6	0	1	0
w_7	0	0	1
w_8	0	0	0

Table 5: Possible Worlds for A, B and C, $f_1(A, w_6)$ in **Bold**, $f_1(A \wedge B, w_6)$ Underlined

The constraints on selection functions listed in Table 5 are partly motivated by our intuitive concept of similarity, but also by logical considerations. Success enforces that the set of most similar p-worlds to w are in fact p-worlds. But, the other constraints correspond to certain logical validities — see Starr (2019) for a thorough discussion of this. For the purposes of this chapter, only strong centering will matter, as it ensures that similarity conditionals validate modus ponens.[13]

Material antecedent and material negation are invalid for the similarity conditional for the same reasons that they were invalid for the strict conditional. The falsity of A in w does not ensure that A-worlds most similar to w are B-worlds. So $\neg A \nvDash A > B$. $A > B$ can be false in w when A is false in w, but in one of the A-worlds most similar to w, B is false. So $\neg(A > B) \nvDash A$.

The context sensitivity of conditionals can be captured on similarity analyses by highlighting the fact that judgements of similarity are themselves context dependent (Lewis 1973). As Lewis (1973: 67) details, different contexts can make salient different properties of the things we are talking about, and this impacts what counts as a similar world to our own. This is illustrated With the pair in (9), discussed back in §2.

(9) a. If Caesar had been in charge [in Korea], he would have used the atom bomb.
b. If Caesar had been in charge [in Korea], he would have used catapults.

Consider a context where Caesar's brutality is made salient. It will then be held fixed when determining which worlds where Caesar was in charge in Korea count as most similar to our own. As a result, (9-a) will come out true and (9-b) false. Other contexts will have the opposite effect. Quine's (1982: Ch.3) claim that there is no fact of the matter whether these counterfactuals are true comes from failing to embed them in natural contexts. Subsequent work such as K. Lewis (2018) and Ippolito (2016) develops this idea.

Many similarity theorists explicitly limited the analysis to counterfactuals (e.g. Lewis 1973). But Stalnaker (1975) applied the analysis to indicative conditionals by saying that indicatives and counterfactuals differ in how they are context sensitive. Recall that indicatives and counterfactuals can differ in truth-value. Setting aside conspiracy theories, (3) is true and (4) false.

(3) If Oswald didn't kill Kennedy, someone else did.

(4) If Oswald hadn't killed Kennedy, someone else would've.

[13] In fact, weak centering suffices: $w \in f(w, p)$ if $w \in p$.

Stalnaker (1975) explains this contrast in terms of a general account of how assertion works. The mutual assumptions of the conversationalists can be modeled a set of worlds c called the *context set* — the set of worlds compatible with what everyone is assuming, and with what everyone is assuming that everyone is assuming, and so on. When a proposition p is asserted, and accepted, then the new context set is $c \cap p$ — the set of all worlds in the previous context set and compatible with p.

Stalnaker (1975) Analysis of Indicatives

For an indicative conditional $A > B$, if $w \in c$, then $f(w, A) \subseteq c$.
When evaluated in a world compatible with the context set, the most similar antecedent worlds must also be within the context set.

This predicts (3) to be true, since it requires that its interpretation hold fixed the fact that Kennedy was killed — that proposition is part of the context set against which (3) is asserted. By contrast, (4) does not require that to be held fixed, and so allows the most similar antecedent worlds to be ones where Kennedy was not killed at all. On this analysis, all conditionals are context sensitive, but indicative conditionals are specifically sensitive to the context set. This goes some way in explaining the observation that indicative conditionals are only felicitous when their antecedent has not been explicitly ruled out — recall (1) and (2) from §1.

3.3 Restrictor Analyses

The restrictor analysis of conditionals originates with Lewis (1975) and Kratzer (1981, 1986), and argues for a dramatic change in the logical analysis of conditionals. It has be assumed that conditionals have a logical form like $A \to B$ and that their analysis must proceed by finding the right meaning for $\to$. But restrictor analyses contend that this is wrong. An analysis of conditionals should begin with modal adverbs like *must*, *would*, *might*, *probably* that occur in what is normally thought to be the consequents of conditionals. A conditional like *if Maya sang, Nelson probably danced* should be thought of primarily as a sentence of the form *Probably*(D). All the *if*-clause does is restrict the domain of worlds over which *Probably* quantifies. While *Nelson probably danced* says that Nelson danced in most of the worlds, *if Maya sang, Nelson probably danced* says that Nelson danced in most of the worlds where Maya sang.

Lewis (1975) and Kratzer (1986) argue for a restrictor analysis by observing that no uniform contribution can be assigned to $A \to B$ that captures the different meanings that conditionals have when different modal adverbs occur in the consequent. This argument will not be summarized here because restrictor analyses are not actually competitors to strict and similarity analyses — they simply specify a different form those two analyses can take. As Kratzer (1991: 649) details, material conditionals, strict conditionals and similarity conditionals can all be modeled within a restrictor analysis. Instead, a restrictor analysis provides a resource for extending the empirical coverage of strict and similarity analyses. While the conditionals considered so far involve a necessity modal of some sort, a more general analysis is needed.[14] Further, they provide a more flexible theory that could, in principle, provide an account on which conditionals are sometimes strict and sometimes similarity conditionals. For more on this approach see Kratzer (2012).

3.4 Dynamic Analyses

This chapter has assumed that the meaning of a sentence corresponds to its truth-conditions: the set of worlds in which it is true. Formally, this means that the semantics is specified as a function

[14] Where there is no overt modal, Kratzer (1986) proposes that conditionals involve a covert epistemic necessity modal.

$[\![A]\!]$ that maps sentences A of the formal language to subsets of W (the set of all possible worlds). Logic followed suite: it requires that valid arguments preserve truth. But, there is another dynamic perspective on meaning: it is the characteristic way a sentence changes a context in which it is uttered (Veltman 1996; Heim 1982; Groenendijk and Stokhof 1991; Kamp 1981).

On Veltman's (1996) approach, a dynamic semantics is specified as a function $[A]$ that maps one information state s (a subset to W) to another information state s'. This is written $s[A] = s'$ and it is said the s' is the result of updating s with A. The meaning of A resides in the difference between the two states: one prior to the use of A, and one posterior to the use of A. A express truth-conditional information about worlds by eliminating worlds from s. But it can also express a global property of the information s that is quite different than distinguishing between worlds.

For example, the sentence *Might*(A) tests whether s is consistent with A. If not, s' is reduced to $\varnothing$. But if it is, then s' is left as it is. On this view, *Might*(A) expresses a property of the information state s, without treating that property as something that distinguishes *worlds* in s from each other. Logic follows suite: valid arguments preserve information rather than truth. More specifically, updating with the conclusion after updating with the premises provides no additional information.

Dynamic Validity

$P_1, \ldots, P_n \vDash C$ just in case for any information state s, $s[P_1]\cdots[P_n] = s[P_1]\cdots[P_n][C]$.

Both strict and similarity analyses have drawn inspiration from this dynamic approach.

von Fintel (2001) and Gillies (2007) develop dynamic strict analyses of counterfactuals and argue that they can better explain ordering effects. Among them are *reverse Sobel sequences*, which are simply the reversal of the sequences like (14) presented by Lewis (1973: 10). The important observation is that reversing these sequences is not felicitous:

(16) a. If I had shirked my duty and you had too, harm would have ensued.
b. #If I had shirked my duty, no harm would have ensued.

von Fintel (2001) and Gillies (2007) observe that similarity analyses render sequences like (16) semantically consistent. Their theories predict this infelicity by providing a systematic theory of how counterfactuals update context.[15] These analyses involve richer models of context than Veltman (1996).

Gillies (2009, 2004) develops a dynamic strict analysis of indicative conditionals on which $s[A \to B]$ tests that all the A-worlds in s are C-worlds. If the test is passed, s stays as it is. If it is failed, s is reduced to $\varnothing$. This analysis, and the dynamic definition of validity, navigates a number of tricky issues in conditional logic. Gillies (2004) uses it to diffuse counterexamples to modus ponens (McGee 1985). Willer (2012) extends this solution to counterexamples to modus ponens presented by Kolodny and MacFarlane (2010). Stojnić (2016) integrates a dynamic strict analysis with a theory of discourse coherence and modal anaphora to address counterexamples to modus ponens and modus tollens. Gillies (2009) uses a dynamic strict analysis to counter a number of arguments (e.g. Gibbard 1981; Edgington 1995) which say that it is not possible to assign truth-conditions to indicative conditionals that are stronger than a material conditional, weaker than a classical strict conditional and capture the *import-export equivalence* $A \to (B \to C)$ and $(A \wedge B) \to C$. Non-dynamic similarity theories invalidate import-export, but offer no evidence in its favor.

Starr (2014) proposes that the general meaning of conditionals is that given by Gillies (2009), but that counterfactuals contain an operator in their antecedent that allows them to talk about the most similar antecedent worlds. This operator is argued for on the basis of work in linguistics such as Iatridou (2000). Starr (2014) argues that this dynamic strict analysis of indicatives and

[15] Moss (2012) and Lewis (2018) are similarity analyses that explain this data pragmatically.

dynamic similarity analysis of counterfactuals solves a number of outstanding problems. Willer (2017) has developed a related account, and argues that only a dynamic strict analysis can explain why indicative versions of (14) are felicitous. This remains an active area of research.

3.5 Discussion

Both strict and similarity analyses can overcome the basic shortcomings of the material conditional analysis. They validate modus ponens while invaliding material antecedent and material negation, and are not truth-functional. They also leave room for the context sensitivity of conditionals. The main debate between them is whether non-monotonic inferences like antecedent strengthening should be rendered invalid (similarity analyses), or whether the proposed counterexamples involve changes in context (strict analyses). This is a subtle and ongoing debate. There is, however, a more pressing challenge to both of these analyses.

To explain how a given conditional like (17) expresses a true proposition, a similarity analysis must specify which particular conception of similarity informs it.

(17) If my computer were off, the screen would be blank.

Of course, the strict analysis is in the same position. It cannot predict the truth of (17) without specifying a particular accessibility relation. In turn, the same question arises: on what basis do ordinary speakers determine some worlds to be accessible and others not? Theories like those discussed above do not directly address this question, as they are primarily concerned with the logic of conditionals. But there is a wealth of examples which illustrate that there are systematic generalizations about how sentences are judged to be true, and it is not clear that strict or similarity analyses are well-position to capture this. Readers interested in this next turn in the debate are directed to Starr (2019: §2.5).

4 Conditionals and Probability

Conditional reasoning involves uncertainty and evaluating the consequences of the world being different than it actually is. The study of probability has provide a suite of tools designed for just these purposes. It is no accident, then, that many philosophers, psychologists and computer scientists have drawn on these tools to analyze the meaning and logic of conditionals. The seminal proposal here comes from Adams (1975) which analyzes conditionals in terms of conditional probability. This basic idea, and its development by philosophers (Edgington 1995) and psychologists (Evans and Over 2004), is surveyed in §4.1. Recently, research in AI, philosophy, psychology and linguistics has appealed to a related probabilistic tool: Bayesian networks. Section 4.2 will survey those accounts.

4.1 Conditional Probability

The seminal proposal from Adams (1975) is this:

Adams' Thesis The assertability of *B if A* is proportional to $P(B \mid A)$, where P is a probability function representing the agent's subjective credences.

Probabilities are real numbers between 0 and 1 assigned to propositional variables $A, B, C, \ldots$. Adams takes these probabilities to reflect an agent's subjective credence, e.g., $P(A) = 0.6$ reflects that they think A is slightly more likely than not to be true.[16] $P(B \mid A)$ is the credence in B conditional on A being true and is defined as follows:

Conditional Probability

$$P(B \mid A) = \frac{P(A \wedge B)}{P(B)}$$

Surprisingly, the logic of conditionals based on conditional probability developed by Adams (1975) is more or less the same as the similarity analyses pursued by Stalnaker (1968) and Lewis (1973). In particular, it invalidates antecedent-monotonic patterns, along with material negation and material antecedent. Its major differences from similarity analyses comes in its account of context-sensitivity and non-truth-functionality. This can be illustrated with examples discussed earlier.

Recall the context of (8). You know a standard die has been tossed, but you do not know which side has landed face up. (8-a) seems like a good assertion to make, while (8-b) does not.

(8) a. If the die came up 2, it came up even.
b. If the die came up 1, it came up even.

When you get to see that the die came up 3, it still seems right to say that (8-a) *was* a justified assertion even if you would not now assert that conditional.[17] The same goes for (8-b). Both intuitions are captured on a conditional probability analysis. The conditional probability of the die coming of even given that it came up 2 is 1, while the conditional probability of the die coming of even given that it came up 1 is 0. (8) was an example of non-truth-functionality, and the conditional probability analysis captures this: instead of the truth of a conditional being determined by the truth of its parts, the probability of a conditional is being determined by the probability of its parts. As it turns out, this view is even more radical. A conditional probability $P(B \mid A)$ cannot be modeled as the probability that a conditional proposition $B \mid A$ is true (Lewis 1976). So $P(B \mid A)$ should not be thought of as the probability that a proposition is true. For this reason, many philosophers articulate a conditional probability analysis as holding that conditionals do not express truth-evaluable propositions, but merely a ratio of credences in truth-evaluable propositions (Edgington 1995).

The context-sensitivity of conditionals is reflected in the core idea of the conditional probability analysis: conditionals express credences. As those credences change, so will the ratios between them. Consider again (10), and it's context: we have mutually established that a die in our possession has 3 on every side except for one, which has 2. The die has been tossed, but we do not know how it came up. (10) seems like a justified assertion.

(10) If the die came up even, it came up 2.

Indeed, the conditional probability of the die coming up 2, given that it came up even is 1. In this context our credence tht the die came up any even number other than 2 is 0, so $P(\mathit{Even} \wedge \mathit{Two}) = P(\mathit{Two})$ and $\frac{P(\mathit{Two})}{P(\mathit{Two})} = 1$. But, when we believe the die to be a standard one, the assertion of (10) will be quite unjustified. Believing the die to be standard amounts to these credences: $P(\mathit{Even}) = \frac{1}{2}$, $P(\mathit{Two}) = \frac{1}{6}$ and $P(\mathit{Even} \wedge \mathit{Two}) = \frac{1}{6}$. So $P(\mathit{Even} \mid \mathit{Two}) = \frac{1/6}{1/2} = \frac{1}{3}$.

Since conditional probabilities are only defined when the antecedent is assigned a non-zero probability, Adams' Thesis is of limited use for counterfactuals. Further, indicative and counterfactual pairs often differ in their assertability, e.g. (3) and (4). To address this, Adams (1976) proposed the *prior probability* analysis of counterfactuals:

[16] Probabilities are taken to obey the Kolmolgorov Axioms. See Hájek (2017) for details.

[17] As noted in §1, indicatives are infelicitous when their antecedent is known to be false. This is expected on Adams' analysis, since $P(B \mid A)$ is undefined when $P(A) = 0$.

Adams' Prior Probability Analysis The assertability of *B would have been, if A had been* is proportional to $P_0(B \mid A)$, where P_0 is the agent's credence prior to learning that A is false.

Consider a counterfactual variant of (10) uttered in the context where the die is known to have 3 on five sides and 2 on one, and we have just learned that the die came up 3.

(18) If the die had come up even, it would have came up 2.

Prior to learning that the die came up 3, and so didn't come up even, $P(\textit{Even} \mid \textit{Two}) = 1$, as just discussed. So the Prior Probability analysis correctly predicts this utterance of (18) to be a perfect assertion. As with the indicative, this would change if we believe the die to be 12-sided instead.

There are other cases, however, where Adams' Prior Probability Analysis makes incorrect predictions. Edgington (2004: 21) discusses these cases and concludes that the analysis requires an amendment: P_0 may also reflect any facts the agent learns after they learn that the antecedent is false, provided that those facts are causally independent of the antecedent. Kvart (1986) integrates causal information into a different objective conditional probability analysis. More recently, this focus on causality and probability has inspired a different approach to counterfactuals. Many using probabilistic tools now favor modeling causal information in terms of Bayesian networks (Pearl 2009). It is then possible to formulate a semantics for counterfactuals directly in terms of Bayesian networks. As §4.2 will explain, Bayesian networks have significant advantages to a standard probabilistic representation when trying to formulate a computationally tractable representation of an agent's knowledge about the world.

4.2 Bayesian Networks

When an agent's credence in B is the same as their credence in B given A and B given $\neg A$, B is probabilistically independent of A:

Probabilistic Independence B is probabilistically independent of A just in case $P(B) = P(B \mid A) = P(B \mid \neg A)$.

Bayesian networks are built on the mathematical insight that it is possible to represent an agent's credences by representing only the conditional probabilities of dependent variables, and the probabilities of independent variables. This mathematical insight has taken to be of great importance to artificial intelligence and cognitive science. It makes probabilistic representations of agents beliefs computationally tractable.[18] But, it also stores immensely useful information. It facilitates counterfactual reasoning, reasoning about actions, and explanatory reasoning.

An agent's knowledge about a system containing 8 variables could be represented by the directed acyclic graph and system of structural equations between those variables in Figure 1. While the arrows mark relations of probabilistic dependence, the equations characterize the nature of the dependence, e.g. '$H = F \vee G$' means that the value of H is determined by the value of $F \vee G$ (but not vice versa). Consider just the three rightmost nodes of Figure 1. The are an appropriate representation for an agent who has credences about three propositions, and their probabilistic dependencies correspond to the indicative conditionals (19-a) and (19-b). Let us further suppose they have the unconditional credence corresponding to (19-c).

(19) a. If both Fran and Greta attend, Harriet attends.

[18] A complete description of an agent's credences involves a joint probability distribution over all Boolean combinations of the variables. For a system with 8 variables this requires storing $2^8 = 256$ probability values, while the Bayesian network would require only 18 — one conditional probability for each boolean combination of the parent variables, and one for each of independent variables. See Sloman (2005: Ch.4) and Pearl (2009: Ch.1) for details.

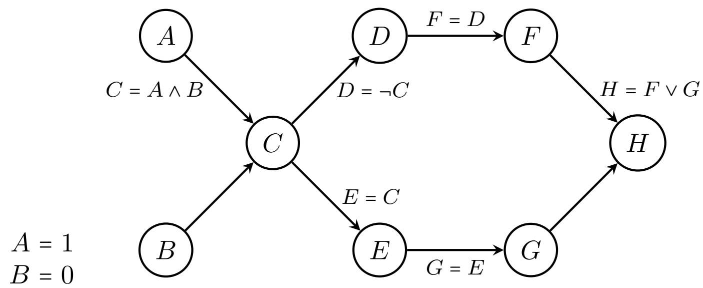

Figure 1: Bayesian Network and Structural Equations

b. If either Fran or Greta don't attend, Harriet won't attend.
c. Fran attended, Greta did not, and so Harriet did not.
d. If Greta had attended, Harriet would have attended.

The counterfactual (19-d) seems true in this scenario. Pearl's (2009: Ch.7) analysis can capture this:

Interventionism Evaluate $G > H$ relative to a Bayesian Network by removing any incoming arrows to G, setting its value to 1, and projecting this change forward through the remaining network. If H is 1 in the resulting network, $G > H$ is true; otherwise it's false.

On this method, one first intervenes on G: remove the arrow coming in to G and the equation $G = E$, and replace it with $G = 1$. One then solves for H using the equation $H = F \wedge G$. Since intervention does not effect the value of F, it remains 1. So, it follows that $H = 1$ and that the counterfactual is true. Pearl (2009: Ch.7) shows that the logic of interventionist counterfactuals is very close to similarity analyses (Lewis 1973; Stalnaker 1968) — hence my use of '>' here.

Unlike similarity analyses, Bayesian networks provide explicit models of the knowledge that makes counterfactuals true. This allows it to better navigate the numerous counterexamples to the similarity analysis surveyed in Starr (2019: §2.5), and provide an explicit theory of how counterfactuals are context-sensitive. To be sure, interventionism has limitations, and faces a number of counterexamples.[19] But there is now a burgeoning interdisciplinary literature refining interventionism (Schulz 2007, 2011; Kaufmann 2013; Lucas and Kemp 2015), and pursuing alternatives also based on Bayesian networks (Hiddleston 2005; Rips 2010). This is a very active area of research.

5 Conclusion

Recent work on the logic of conditionals maintains that they have three key properties: they are non-truth-functional, they are context sensitive, and their antecedents are interpreted non-monotonically. Certain core validities, like modus ponens, and invalidities, like material negation and material antecedent, have been captured alongside these key properties. Possible worlds analyses have come in two basic varieties: strict analyses and similarities analyses. These two varieties have been augmented in various ways using restrictor analyses of modality and dynamic semantics. Accounting

[19]It does not apply to conditionals with logically complex antecedents or consequents. This limitation is addressed by Briggs (2012), who also axiomatizes and compares the resultant logic to Lewis (1973) and Stalnaker (1968) — significantly extending the analysis and results in Pearl (2009: Ch.7). For counterexamples see Hiddleston (2005)and Rips (2010).

for the particular contextual features that fix the truth-conditions of conditionals remains a challenge for these approaches. Probabilistic analyses present a promising option here, but are still very much in development.

References

Adams, E. W. (1975). *The Logic of Conditionals*. Dordrecht: D. Reidel.

Adams, E. W. (1976). Prior probabilities and counterfactual conditionals. In W. Harper and C. Hooker (Eds.), *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, Volume 6a of *The University of Western Ontario Series in Philosophy of Science*, pp. 1–21. Springer Netherlands.

Anderson, A. R. (1951). A note on subjunctive and counterfactual conditionals. *Analysis 12*(2), pp. 35–38.

Arlo-Costa, H. and P. Egré (2016). The logic of conditionals. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016 ed.). Metaphysics Research Lab, Stanford University.

Briggs, R. A. (2012). Interventionist counterfactuals. *Philosophical Studies 160*(1), 139–166.

Declerck, R. and S. Reed (2001). *Conditionals: A Comprehensive Emprical Analysis*, Volume 37 of *Topics in English Linguistics*. New York: De Gruyter Mouton.

Dudman, V. H. (1988). Indicative and subjunctive. *Analysis 48*(3), 113–122.

Edgington, D. (1995). On conditionals. *Mind, New Series 104*(413), 235–329.

Edgington, D. (2004). Counterfactuals and the benefit of hindsight. In P. D. . P. Noordhof (Ed.), *Cause and Chance: Causation in an Indeterministic World*, pp. 12–27. New York: Routledge.

Edgington, D. (2014). Indicative conditionals. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2014 ed.). Metaphysics Research Lab, Stanford University.

Evans, J. S. B. T. and D. Over (2004). *If.* Oxford: Oxford University Press.

Gibbard, A. F. (1981). Two recent theories of conditionals. In W. L. Harper, R. C. Stalnaker, and G. Pearce (Eds.), *Ifs: Conditionals, Beliefs, Decision, Chance, Time*, pp. 211–247. Dordrecht: D. Reidel.

Gillies, A. (2007). Counterfactual scorekeeping. *Linguistics & Philosophy 30*(3), 329–360.

Gillies, A. (2012). Indicative conditionals. In G. Russell and D. G. Fara (Eds.), *The Routledge Companion to Philosophy of Language*, pp. 449–465. New York: Routledge.

Gillies, A. S. (2004). Epistemic conditionals and conditional epistemics. *Noûs 38*(4), 585–616.

Gillies, A. S. (2009). On truth-conditions for “if” (but not quite only “if”). *Philosophical Review 118*(3), 325–349.

Goodman, N. (1947). The problem of counterfactual conditionals. *The Journal of Philosophy 44*, 113–118.

Grice, P. (1989). Indicative conditionals. In *Studies in the Way of Words*, Chapter 4, pp. 58–85. Cambridge, Massachuesetts: Harvard University Press.

Groenendijk, J. and M. Stokhof (1991). Dynamic predicate logic. *Linguistics and Philosophy 14*(1), 39–100.

Hájek, A. (2017). Interpretations of probability. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2017 ed.). Metaphysics Research Lab, Stanford University.

Heim, I. R. (1982). *The Semantics of Definite and Indefinite Noun Phrases*. Ph. D. thesis, Linguistics Department, University of Massachusetts, Amherst, Massachusetts.

Hiddleston, E. (2005). A causal theory of counterfactuals. *Noûs 39*(4), 632–657.

Iatridou, S. (2000). The grammatical ingredients of counterfactuality. *Linguistic Inquiry 31*(2), 231–270.

Ippolito, M. (2016). How similar is similar enough? *Semantics and Pragmatics 9*(6), 1–60.

Kamp, H. (1981). A theory of truth and semantic representation. In J. A. Groenendijk, T. Janssen, and M. Stokhof (Eds.), *Formal Methods in the Study of Language*, pp. 277–322. Dordrecht: Foris.

Kaufmann, S. (2013). Causal premise semantics. *Cognitive Science 37*(6), 1136–1170.

Kolodny, N. and J. MacFarlane (2010). Ifs and oughts. *Journal of Philosophy 107*(3), 115–143.

Kratzer, A. (1981). The notional category of modality. In H.-J. Eikmeyer and H. Rieser (Eds.), *Words, Worlds and Contexts*, pp. 38–74. Berlin: Walter de Gruyter.

Kratzer, A. (1986). Conditionals. In *Proceedings from the 22nd Regional Meeting of the Chicago Linguistic Society*, Chicago, pp. 1–15. University of Chicago.

Kratzer, A. (1991). Modality. In A. von Stechow and D. Wunderlich (Eds.), *Semantics: An International Handbook of Contemporary Research*, pp. 639–650. Berlin: De Gruyter Mouton.

Kratzer, A. (2012). *Modals and Conditionals: New and Revised Perspectives*. New York: Oxford University Press.

Kripke, S. A. (1963). Semantical analysis of modal logic I: Normal modal propositional calculi. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik 9*, 67–96.

Kvart, I. (1986). *A Theory of Counterfactuals*. Indianapolis, IN: Hackett.

Lewis, D. K. (1973). *Counterfactuals*. Cambridge, Massachusetts: Harvard University Press.

Lewis, D. K. (1975). Adverbs of quantification. In E. L. Keenan (Ed.), *Formal Semantics of Natural Language*, pp. 3–15. Cambridge, England: Cambridge University Press.

Lewis, D. K. (1976). Probabilities of conditionals and conditional probabilities. *The Philosophical Review 85*(3), 297–315.

Lewis, K. S. (2018). Counterfactual discourse in context. *Noûs 52*(3), 481–507.

Lucas, C. G. and C. Kemp (2015). An improved probabilistic account of counterfactual reasoning. *Psychological Review 122*(4), 700–734.

McGee, V. (1985). A counterexample to modus ponens. *Journal of Philosophy 82*(9), 462–471.

McKay, T. J. and P. van Inwagen (1977). Counterfactuals with disjunctive antecedents. *Philosophical Studies 31*, 353–356.

Moss, S. (2012). On the pragmatics of counterfactuals. *Noûs 46*(3), 561–586.

Nute, D. (1975, 10). Counterfactuals. *Notre Dame J. Formal Logic 16*(4), 476–482.

Pearl, J. (2009). *Causality: Models, Reasoning, and Inference* (2nd ed.). Cambridge, England: Cambridge University Press.

Peirce, C. S. (1896). The regenerated logic. *The Monist 7*(1), pp. 19–40.

Pollock, J. L. (1976). *Subjunctive Reasoning*. Dordrecht: D. Reidel Publishing Co.

Quine, W. V. O. (1982). *Methods of Logic* (4th ed.). Cambridge, MA: Harvard University Press.

Rips, L. J. (2010). Two causal theories of counterfactual conditionals. *Cognitive Science 34*(2), 175–221.

Schulz, K. (2007). *Minimal Models in Semantics and Pragmatics: Free choice, Exhaustivity, and Conditionals*. Ph. D. thesis, University of Amsterdam: Institute for Logic, Language and Computation, Amsterdam.

Schulz, K. (2011). If you'd wiggled a, then b would've changed. *Synthese 179*, 239–251.

Sloman, S. (2005). *Causal Models: How People Think About the World and Its Alternatives*. New York: OUP.

Stalnaker, R. (1968). A theory of conditionals. In N. Rescher (Ed.), *Studies in Logical Theory*, pp. 98–112. Oxford: Basil Blackwell.

Stalnaker, R. (1975). Indicative conditionals. *Philosophia 5*, 269–286.

Stalnaker, R. C. and R. H. Thomason (1970). A semantic analysis of conditional logic. *Theoria 36*, 23–42.

Starr, W. B. (2014). A uniform theory of conditionals. *Journal of Philosophical Logic 43*(6), 1019–1064.

Starr, W. B. (2019). Counterfactuals. In E. N. Zalta (Ed.), *Stanford Encylcopedia of Philosophy* (Spring 2019 ed.). Metaphysics Research Lab, Stanford University.

Stojnić, U. (2016). One's modus ponens: Modality, coherence and logic. *Philosophy and Phenomenological Research*.

Veltman, F. (1986). Data semantics and the pragmatics of indicative conditionals. In E. C. Traugott, A. ter Meulen, J. S. Reilly, and C. A. Ferguson (Eds.), *On Conditionals*. Cambridge, England: Cambridge University Press.

Veltman, F. (1996). Defaults in update semantics. *Journal of Philosophical Logic 25*(3), 221–261.

von Fintel, K. (1999). The presupposition of subjunctive conditionals. In U. Sauerland and O. Percus (Eds.), *The Interpretive Tract*, Volume MIT Working Papers in Linguistics 25, pp. 29–44. Cambridge, MA: MITWPL.

von Fintel, K. (2001). Counterfactuals in a dynamic context. In M. Kenstowicz (Ed.), *Ken Hale: a Life in Language*, pp. 123–152. Cambridge, Massachusetts: The MIT Press.

von Fintel, K. (2012). Subjunctive conditionals. In G. Russell and D. G. Fara (Eds.), *The Routledge Companion to Philosophy of Language*, pp. 466–477. New York: Routledge.

Warmbrōd, K. (1981). An indexical theory of conditionals. *Dialogue, Canadian Philosophical Review 20*(4), 644–664.

Willer, M. (2012). A remark on iffy oughts. *Journal of Philosophy 109*(7), 449–461.

Willer, M. (2017). Lessons from sobel sequences. *Semantics and Pragmatics 10*(4).