11

CONDITIONALS

R. A. Briggs

Conditionals are sentences that propose a scenario (which may or may not be the actual scenario), then go on to say something about what would happen in that scenario.[1] In English, they are typically expressed by 'if... then...' statements. Examples of conditionals include:

1. If the Axiom of Choice is true, then every set can be well ordered.
2. You will probably get lung cancer if you smoke.
3. If the syrup forms a soft ball when you drop it into cold water, then it is between 112 and 115 degrees Celsius.
4. If kangaroos had no tails, they would topple over.
5. When I'm queen, you will be sorry.

In general, a conditional is formed from two smaller statements: an *antecedent* (the supposition that typically comes directly after 'if') and a *consequent* (the statement that typically comes later in the sentence, and is sometimes preceded by 'then'). In the above examples, the antecedents are:

1. The Axiom of Choice is true.
2. You smoke.
3. The syrup forms a soft ball when you drop it into cold water.
4. Kangaroos had no tails. (Or perhaps: Kangaroos have no tails.)
5. I'm queen.

while the consequents are:

1. Every set can be well ordered.
2. You will probably get lung cancer. (Or perhaps: You will get lung cancer.)
3. The syrup is between 112 and 115 degrees Celsius.
4. Kangaroos would topple over. (Or perhaps: Kangaroos topple over.)
5. You will be sorry.

1 I take this framing, which emphasizes the contents of conditionals rather than their grammatical form, from (von Fintel, 2011).

1 WHY CARE ABOUT CONDITIONALS?

Conditionals are useful for a variety of everyday tasks, including decision making, prediction, explanation, and imagination.

When making a decision, you should aim to choose an act such that, if you (were to) perform it, a good outcome is (or would be) likely to result. Decision theory codifies this intuition in formal terms, and often makes explicit use of conditionals (Gibbard & Harper, 1981; Vinci, 1988; Bradley, 2000; Cantwell, 2013).

Conditionals are also useful for deriving predictions and explanations from theoretical models. If I am not sure which model of climate change to accept, I can use conditionals to reason about how much the earth's temperature will increase if each of the models under consideration is true. To check whether a model explains the data I have already observed, I can use conditionals to check whether, if a given model is true, my data should be expected. (For a defense of conditionals in scientific explanation, see Woodward, 2004; for a defense of conditionals in historical explanation, see Reiss, 2009, and Nolan, 2013.)

Children's pretend play is both developmentally important, and closely related to reasoning with conditionals. Amsel and Smalley (2000), Dias and Harris (1990), Gopnik (2009), Harris (2000), Lillard (2001), and K. Walton (1990) argue that children's pretense (for example, pretending a banana is a telephone), involves constructing an alternative scenario to what is known or believed to be true, and then reasoning about what would happen in that scenario. While children can express their thoughts about pretend scenarios without the explicit use of conditionals, conditionals are particularly well suited to expressing these thoughts. Weisberg and Gopnik (2013) argue that the ability to reason about non-actual scenarios is crucial to learning from and planning for the actual world, since it enables children to generate and compare a range of alternative models of reality. Krzyzanowska (2013) argues that the mechanism that lets children evaluate conditionals is the same as the one that lets them attribute false beliefs to others.

In addition to playing a crucial role in everyday reasoning and cognitive development, conditionals do work in philosophical analyses of a variety of concepts. Any philosophical idea that relies on the notion of dependence is ripe for a conditional analysis: to say that one thing e depends on a second thing c is arguably to say that if c is one way, then e is some corresponding way, and if c is a different way, then e is a correspondingly different way. Conditionals famously appear in analyses of causation (see Menzies, 2014, and Collins, Hall, and Paul, 2004, for overviews), dispositions (Prior, Pargetter, & Jackson, 1982; Choi, 2006, 2009), knowledge (Nozick, 1981; Sosa, 1999), and freedom (Moore, 1912; Ayer, 1954).

Finally, conditionals figure in several common patterns of reasoning, to which we now turn.

2 COMMON PATTERNS OF REASONING

The following argument forms look compelling in ordinary, natural-language arguments (though we will see that all of them have putative counterexamples). Different formal theories of conditionals yield different verdicts about which are valid.

2.1 *Modus Ponens*

Modus ponens is the inference form:

1. If A, then C.
2. A.

$\therefore$ C.

Modus ponens is one of the most central—arguably the most central—of the inference forms involving conditionals. Bobzien (2002) traces its roots back to Aristotle's hypothetical syllogisms, and through the logic of the Peripatetics and antiquity. Gillon (2011) notes that modus ponens was a common inference pattern in Pre-Classical Indian philosophy, and quotes a representative argument in which the third-century Buddhist logician Moggaliputta Tissa explicitly notes the inconsistency of simultaneously believing 'if A, then C', 'A', and 'not C'. Ryle (1950) even advances a theory of conditionals based entirely on their ability to license modus ponens: an utterance of 'if A then C' is an 'inference ticket' that allows one to move from the premise A to the conclusion C.

Despite its perennial popularity, there are apparent counterexamples to modus ponens. One sort (McGee, 1985) involves nested conditionals. Suppose you see a fisherman with something caught in his net. You are almost sure it is a fish, but the next likeliest option is that it is a frog. McGee argues that you should accept the premises of the following argument, but not the conclusion (since, if the animal has lungs, then it is not a fish but a frog).

1. If that is a fish, then if it has lungs, it's a lungfish.
2. That is a fish.

$\therefore$ If it has lungs, it's a lungfish.

Another type of apparent counterexample (Kolodny & MacFarlane, 2010; Darwall, 1983) involves 'ought's or 'should's. Consider this variant of Darwall's example.

1. If you want to hurt my feelings, you should make fun of the way my ears stick out.
2. You want to hurt my feelings.

∴ Therefore, you should make fun of the way my ears stick out.

Even if you do want to hurt my feelings, you shouldn't make fun of the way my ears stick out, because it's wrong to hurt my feelings. Dowell (2011), and Lauer and Condoravdi (2014) object to the Darwall example (and other, related examples) on the grounds that they equivocate on different meanings of 'should'.

Yet another type of apparent counterexample to modus ponens, discussed by D. Walton (2001), involves defeasible inferences, like the famous Tweety Bird example from cognitive science (Brewka, 1991).

1. If Tweety is a bird, then Tweety flies.
2. Tweety is a bird.

∴ Tweety flies.

The first premise of the Tweety bird argument says that there is a defeasible connection between being a bird and flying—one that can be overridden by extra information, e.g., that Tweety is a penguin. Thus, the premises are true, and the conclusion false, in the case where Tweety is a penguin.

2.2 *Modus Tollens*

Modus tollens is the inference form:

1. If A, then C.
2. Not C.

∴ Not A.

Modus ponens and modus tollens seem to have originated together (see Bobzien, 2002, and Gillon, 2011), and are closely related. Both inferences posit a three-way inconsistency between 'if A, then C', 'A' and 'not C'. Affirm two of these inconsistent claims, and you'll have to deny the third.

Yalcin (2012a) presents a putative counterexample to modus tollens. Consider an urn that contains 100 marbles—some red, some blue, some big, and some small—in the following proportions.

	blue	red
big	10	30
small	50	10

A marble is chosen at random and placed under a cup; no other information about the situation is available.

In Yalcin's scenario, it is reasonable to accept the premises, but not the conclusion, of this instance of modus tollens.

1. If the marble is big then it's likely red.
2. The marble is not likely red.

∴ The marble is not big.

2.3 *Conditional Proof*

Conditional proof (sometimes called the *deduction theorem* in formal logic) lets us establish conditional conclusions without relying on any conditional assumptions. Suppose that an argument from the premises X and A to the conclusion C is valid. Then conditional proof lets us conclude that the argument from X to 'if A, then C' is valid. (Unlike modus ponens and modus tollens, which let us reason from the truth of some propositions to the truth of another proposition, conditional proof lets us reason from the validity of one argument to the validity of another.)

Stalnaker (1975) gives an argument that can easily be worked into a counterexample to conditional proof (though he does not present it that way). The following argument is valid, since in classical logic, anything follows from a contradiction:

1. The butler did it.
2. The butler didn't do it.

∴ The gardener did it.

But the following argument is not valid:

1. The butler did it.

∴ If the butler didn't do it, then the gardener did it.

Although conditional proof in its full generality looks implausible, a restricted version is more appealing: if A all by itself entails C, then 'if A, then C' is a truth of logic. (Koons (2014) makes a similar suggestion about conditional proof in nonmonotonic logic.)

2.4 *Transitivity, Contraposition, and Strengthening the Antecedent*

Transitivity is the inference form:

1. If A, then B.
2. If B, then C.

∴ If A, then C.

Contraposition is:

1. If A, then C.

∴ If not C, then not A.

And strengthening the antecedent is:

1. If A, then C.

∴ If A and B, then C.

All three inference forms seem to fail for ordinary conditionals in English. For transitivity, we have the following counterexample (Stalnaker, 1968, p. 106):

1. If J. Edgar Hoover had been born a Russian, then he would have been a communist.
2. If J. Edgar Hoover had been a communist, then he would have been a traitor.

∴ If J. Edgar Hoover had been born a Russian, then he would have been a traitor.

For contraposition, we have the following counterexample (adapted from Adams, 1988):

1. If it rains, then it does not rain hard.

∴ If it rains hard, then it does not rain.

And for strengthening the antecedent, we have the following counterexample (Stalnaker, 1968, p. 106):

1. If this match were struck, then it would light.

∴ Therefore, if this match had been soaked in water overnight and it were struck, then it would light.

Not everyone accepts these putative counterexamples as genuine. Brogaard and Salerno (2008) argue that the meaning of a conditional depends partly on a contextually determined set of relevant possible worlds. They claim that the putative counterexamples involve a context shift between the premises and the conclusion, but in any fixed context, the arguments are valid.

Von Fintel (2001), Gillies (2007), and Williams (2008) cite linguistic evidence in support of the context shift hypotheses: changing the order of the premises and conclusions in the counterexample arguments changes whether they seem true or false. Counterexamples to antecedent strengthening are closely related to so-called *Sobel sequences* (named for Sobel 1970). A Sobel sequence consists of two sentences of the following form (Gillies, 2007).

(a) If Sophie had gone to the New York Mets Parade, she would have seen Pedro Martínez.

(b) But if Sophie had gone to the New York Mets Parade and gotten stuck behind a tall person, she would not have seen Pedro Martínez.

It seems perfectly reasonable to assert (a) followed by (b). But once someone has asserted (b), an assertion of (a) seems inappropriate—after all, if Sophie had gone to the parade, who's to say she would not have gotten stuck behind a tall person?

Fintel, Gillies, and Williams claim that Sobel sequences involve a context shift: once someone asserts (b), the context changes to make (a) false, but (a) and (b) are never true in the same context. Moss (2012) proposes an alternative explanation: once (b) has been asserted, (a) might be true, but is no longer known, since asserting (b) changes the standards a belief must meet in order to count as knowledge.

2.5 *Simplification of Disjunctive Antecedents*

Simplification of disjunctive antecedents ('simplification' for short; Nute, 1975) is the argument form:

1. If A or B, then C.

$\therefore$ If A, then C.

Simplification seems appealing on its face: surely, to say that the bus will be late if it rains or snows is to say that the bus will be late if it rains, and the bus will be late if it snows.

However, one can easily generate counterexamples by substituting the same sentence for B and C. Suppose I have enough money to visit either Disneyland or Graceland, but not enough to visit both. Then the premise of the following argument is true, while its conclusion is false.

1. If I visit Disneyland or I visit Graceland, then I'll visit Graceland.

$\therefore$ If I visit Disneyland, then I'll visit Graceland.

Counterexamples to strengthening the antecedent can be used to generate counterexamples to simplification (Fine, 1975). Suppose we have both of the following:

1. If A, then C.
2. Not: if A and B, then C.

A is logically equivalent to [(A and B) or (A and not B)], so by 1, we have:

3. If [(A and B) or (A and not B)], then C.

But by simplification, the truth of 3 would have to entail the falsity of 2.

So there is a three-way tension between the validity of simplification, the invalidity of strengthening the antecedent, and the substitution of

logical equivalents. All three ways out of the puzzle are represented in the literature: Loewer (1976) and Mckay and Inwagen (1977) reject simplification; defenders of strict conditional accounts (Section 4.1) accept strengthening the antecedent; and Nute (1975) and Alonso-Ovalle (2009) reject substitution.

3 THE INDICATIVE/COUNTERFACTUAL DISTINCTION

Conditionals in English can be divided into two categories, exemplified by the following pair of sentences (Adams, 1970):

(DD). If Oswald did not shoot Kennedy, then someone else did.

(HW). If Oswald had not shot Kennedy, then someone else would have.

Although (DD) and (HW) are built up from the same antecedent and consequent, they mean different things. (DD) would be acceptable to most people familiar with US history: Kennedy was shot, so someone must have shot him—if not Oswald, then someone else. But (HW) is more controversial; it is accepted by conspiracy theorists, but rejected by those who believe that Oswald acted alone. Sentences like (DD) are called *indicative*; sentences like (HW) are called *counterfactual* (or sometimes *subjunctive*).

It's not clear how to classify conditionals whose antecedents concern the future. Consider the following sentence, as uttered by a conspirator before the Kennedy assassination.

(DW). If Oswald does not shoot Kennedy, then someone else will.

Dudman (1983, 1984) and Bennett (1988) argue that future-tensed conditionals like (DW) belong with counterfactuals like (HW); Bennett (2003, 2001; yes the same Bennett!) argues that they belong with indicatives like (DD); Edgington (1995) argues that there exist distinct categories of future-tensed indicatives and future-tensed counterfactuals.

Philosophers also disagree about the precise relationship between indicatives and counterfactuals. Some favor what Bennett (2003) calls 'Y-shaped analyses', which first explain what is common to indicatives and counterfactuals, and then bifurcate to explain how this common core can produce two different kinds of conditionals. Others (notably Gibbard, 1981, and Bennett, 2003) argue that we need completely separate theories of indicatives and counterfactuals—that there is no interesting core shared by both.

In what follows, I will write '$A \mathrel{\square\!\!\rightarrow} C$' to indicate a counterfactual conditional; '$A \rightarrow C$' to abbreviate an indicative conditional; and 'if A, then C' where I wish to remain neutral. I turn now to a popular class of theories, typically aimed at explaining counterfactual conditionals, but sometimes extended to cover indicatives.

4 SELECTION FUNCTIONS

One way to give a theory of conditionals is to spell out their *truth conditions*, i.e., the circumstances under which they are true. Formally, philosophers represent the truth conditions of a sentence as a function from possible worlds (i.e., ways the world might be) to truth values. Fully specifying the truth conditions for every conditional would be too tall an order: to understand the truth conditions for 'if ontogeny recapitulates phylogeny, then snakes develop vestigial legs', we would have to understand the truth conditions of 'ontogeny recapitulates phylogeny' and 'snakes develop vestigial legs', and that job falls outside the scope of a theory of conditionals. So theories of conditionals adopt a more modest aim: to give a recipe for deriving the truth conditions for 'if A, then C' from the truth conditions of (arbitrary) A and C.

The concept of a selection function (Stalnaker, 1968) provides a way of assigning truth conditions to a conditional based on the truth conditions of its antecedent and consequent. The basic idea is that, to evaluate 'if A, then C', we should first consider a set of *selected* possible worlds where A is true. (Henceforth, I will use 'A-worlds' as shorthand for 'worlds where A is true'.) Intuitively, the selected worlds represent ways the actual world might be if A were true. We then check whether, at all the selected worlds, C is true. If so, then the counterfactual conditional 'if A, then C' is true at the actual world; otherwise, it is false at the actual world.

More formally, we can model this process in terms of a selection function f that maps ordered pairs consisting of a possible world and a proposition onto sets of possible worlds. 'If A, then C' is true at a possible world w if and only if C is true at every world in $f(A, w)$. Different ways of interpreting the selection function yield different theories of conditionals.

4.1 *Strict Conditionals*

One natural way to interpret the selection function is to check *all* possible A-worlds, and say that 'if A, then C' is true at world w just in case C is true at all of them. (Since what is possible may depend on what is actual, the truth value of the conditional may vary from world to world.) This approach yields the *strict conditional* interpretation of the selection function, first developed by C. Lewis (1918). The strict conditional approach classifies transitivity, contraposition, and antecedent-strengthening as valid—which its opponents claim is a mistake (see D. Lewis, 1973a, pp. 4–12).

The strict conditional interpretation also gives questionable results about which counterfactuals are true. If I were to leap out of the second-story window of my office, I would hurt myself—but the strict conditional account says this is not so. There are possible worlds where I leap out the

second-story window and remain unharmed: some where there is a safety net underneath the window, some where I am thoroughly ensconced in protective bubble wrap, some where my body is much less fragile than ordinary human bodies, some where the Earth's gravitational field is weak... but none of them is the sort of world that would result, if I were to leap out the second-story window. Hájek (manuscript) sums up the problem this way: on the strict conditional interpretation, most counterfactuals are false.[2]

4.2 *Closest Worlds*

An alternative to the strict conditional approach, typically used for counterfactuals, defines the selection function in terms of similarity among possible worlds. For every world w, we can rank worlds from most similar to w ('closest') to least similar ('farthest away'). D. Lewis (1973a) holds that every such ranking is a *total preorder*: two worlds can be equally similar to w, but they must be comparable, so that either they are equally similar or one is more similar than the other. (Stalnaker, 1968, discusses the special case of the logic where no two worlds are equally close to a given world; Pollock, 1976, discusses a generalization where worlds may be incomparable in terms of closeness.) $A \boxrightarrow C$ is true at w just in case C is true at all the A-worlds that are most similar to w.

Formally, the closest-worlds interpretation can be modeled using a system of 'spheres'—sets of worlds such that every world in the set is closer to w than every world outside it (D. Lewis, 1973a). Then $f(A, w)$ is the intersection of the set of A-worlds with the smallest sphere containing at least one A-world.[3]

Unlike the strict conditional interpretation, the closest-worlds interpretation of the selection function can explain why transitivity, contraposition, and antecedent-strengthening seem invalid. On the closest-worlds interpretation, they *are* invalid, and we can use diagrams (adapted from D. Lewis, 1973a) to illustrate why.

To see why transitivity is invalid, consider a system of spheres model centered on a particular world w, depicted in Figure 1a. (Worlds are points in the diagram, and spheres are concentric circles.) The A-worlds are the points inside the shape labeled A, the B-worlds are the points inside

2 Hájek argues that the problem extends beyond strict conditional accounts; it also affects the closest-worlds account in Section 4.2. K. S. Lewis (2015) argues that we can save the closest-worlds account by ignoring worlds that are deemed irrelevant by a contextually-determined standard of relevance.

3 Some technical difficulties arise when there is no smallest sphere containing at least one A-world, but only a limitless sequence of ever-smaller spheres; see D. Lewis (1973b, pp. 424–425); Stalnaker (1981, pp. 96–99); Warmbrod (1982); and Díez (2015) for discussion.

the shape labeled B, and the C-worlds are the points inside the shape labeled C. All the closest A-worlds to w are B-worlds, and all the closest B-worlds are C-worlds; yet none of the closest A-worlds are C-worlds. Figure 1b shows a counterexample to contraposition, and Figure 1c shows a counterexample to antecedent strengthening.

Defenders of the closest-worlds theory have the burden of spelling out what 'closeness' amounts to. D. Lewis (1973a) claims that closeness is based on similarity among worlds: to say that one world is closer to w than another is to say that the first world is more similar to w than the second. But Fine (1975) presents an example where greater similarity does not make for greater closeness. (I have taken a few liberties with the details of the example.)

On September 26, 1983, at the height of the Cold War, a Soviet early-warning system went off, falsely reporting that missiles had been launched at Russia from the US (Aksenov, 2013). The officer who saw the alarm, Stanislav Petrov, did not report it to his superiors, and so Russia did not launch missiles in retaliation. The following conditional seems true:

PETROV. If Petrov had informed his superiors at the time of the false alarm, then there would have been a nuclear war.

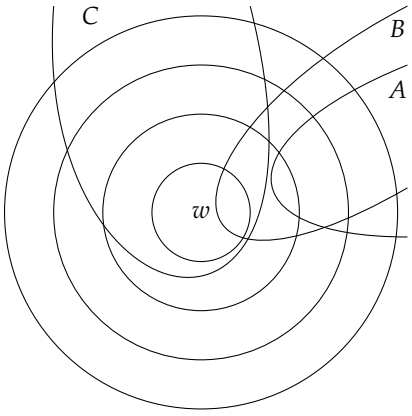
After all, Petrov's superiors were poised to launch the missiles in the event of an attack, and it seems that the phone lines and missile system were in working order. The only missing ingredient was the report from Petrov.

But among the worlds where Petrov informs his superiors at the time of the false alarm, those where the Soviet missile launch is prevented by a happy accident—incompetence by Petrov's superiors, or a broken telephone, or a malfunction of the Soviet missile system—are more similar to the actual world than those where the launch goes through. Worlds where the missile launch is prevented by a happy accident agree with the actual world about the total number of nuclear wars in the 20th Century—surely a more important dimension of similarity than the functioning or malfunctioning of one measly telephone line.[4]

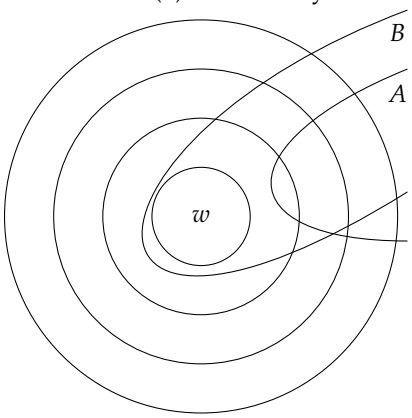
4.3 *Past Predominance*

To handle the PETROV example, a natural thought goes, we need an account of the selection function that treats the past differently from the future. When Petrov made his choice, the missile launch system was already in

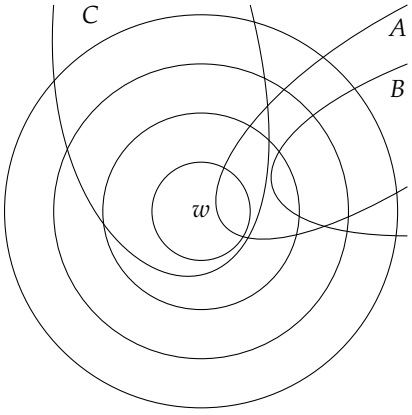
4 Defenders of the closest-worlds interpretation reply that we should understand 'similarity' so that agreeing about the total number of nuclear wars in the 20th Century does not make for greater similarity than agreeing about the functioning or malfunctioning of one measly telephone line; see D. Lewis (1979) and Arregui (2009).



(a) Transitivity



(b) Contraposition



(c) Strengthening the antecedent

Figure 1: Counterexamples to transitivity, contraposition, and strengthening the antecedent in the closest-worlds framework

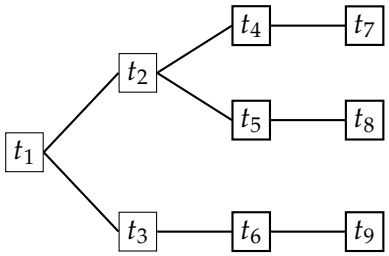


Figure 2: A model of branching time

working order—but it was not yet determined whether there would be a war.

Thomason and Gupta (1980) propose an account of the selection function that takes seriously the past-future asymmetry. They model the universe using branching time, where each moment has only one possible past, but multiple possible futures. (Cross, 1990, shows that the assumption of branching time is dispensable; past predominance can also be modeled using ordinary possible worlds.) Figure 2 depicts such a model. The nodes t_1, t_2, ..., t_9 are moments. Paths through the tree—in this example, $\{t_1, t_2, t_4, t_7\}$, $\{t_1, t_2, t_5, t_8\}$, and $\{t_1, t_3, t_6, t_9\}$—are called *histories*.

We can think of each possible world as containing information about which moment is present, as well as information about which history is actual. (On this way of understanding the model, even when the present moment has more than one possible future, there is a fact of the matter about which future will occur.)

Thomason and Gupta adopt a *past predominance* principle, which says that if a world is in $f(A, w)$, then it must diverge from w as late as possible—there can be no other A-world whose history overlaps w for a longer span than $f(A, w)$.[5]

The past predominance view can explain the PETROV example. Consider the following interpretation of our diagram: the actual history is $\{t_1, t_2, t_4, t_7\}$. At t_1, it is not yet settled whether the early warning system goes off. The early warning system goes off t_2, and Petrov must decide what to do. (At t_3, which belongs to an alternative history, there is never any alarm.) At t_4, Petrov decides not to notify his superiors, and so at t_7, there is no nuclear war. (At t_5, which belongs to another alternative history, Petrov decides to notify his superiors, and a nuclear war ensues at t_8.)

Now consider the conditional PETROV, as uttered at t_7. Its antecedent is false at the actual world, which has the history $\{t_1, t_2, t_4, t_7\}$. The closest

5 For technical reasons, Thomason and Gupta also assume that $f(A, w)$ is a singleton set, and posit that each world contains a *choice function*, which specifies not just what the future will be like, but what the future would have been like had the past gone differently. I pass over the details.

worlds where its antecedent is true must have the history $\{t_1, t_2, t_5, t_8\}$, which diverges from the actual world's history at the last possible moment yet still makes the antecedent true. Since the actual present moment is t_7, it seems reasonable to select t_8 as the present moment at all the closest worlds. Since there is a nuclear war at t_8, the consequent of PETROV is true at all the the closest worlds; hence PETROV is true at the actual world.

4.4 *Causal Models*

A class of examples called *Morgenbesser cases* (Slote, 1978, 27n) suggest that the selection function should respect causal as well as temporal constraints. Edgington (2004) gives a representative Morgenbesser case.

Our heroine misses a flight to Paris due to a car breakdown. She complains to the repairman: 'If I had caught the plane, I would have been halfway to Paris by now!' But he corrects her: 'I was listening to the radio. It crashed. If you had caught that plane, you would be dead by now.'

The repairman claims that the following counterfactual is true.

LETHAL. If the heroine had caught that plane, she would be dead by now.

He is right. It's not clear that past predominance can explain why he's right: the plane crash occurs after our heroine would have made her flight.[6] What matters is that the plane crash is causally independent of whether she makes her flight. This is why, when assessing what would have happened if our heroine had made her flight, we should hold the plane crash fixed.

Pearl (2009) proposes a causal theory of counterfactuals that accounts for Morgenbesser cases. His theory relies on the concept of a *causal model*, consisting of a set of *variables*, which represent what circumscribed parts of the world are like, and a set of *structural equations*, which represent direct causal links between variables. Each variable is assigned an *actual value*; we can think of variables as questions about parts of the world, their possible values as possible answers to those questions, and their actual values as the correct answers in the actual world. Note that although I introduced selection semantics as a recipe for assigning truth values to conditionals at worlds, Pearl's theory is a recipe for assigning truth values to conditionals at model-valuation pairs.[7]

6 But see Phillips (2007) for an argument that past predominance *can* provide an adequate explanation.

7 Pearl's theory can be understood as a version of the situation semantics defended by Barwise and Perry (1981). Instead of assigning truth values to propositions at worlds, it assigns truth values to propositions at situations, which represent ways that circumscribed parts of the world could be.

We can understand Pearl's theory by first building a causal model of Edgington's plane example, then using the model to evaluate the conditional LETHAL. The model will include the following variables.

$$\textsc{car} = \begin{cases} 1 & \text{if the car is working,} \\ 0 & \text{otherwise.} \end{cases}$$

$$\textsc{catch} = \begin{cases} 1 & \text{if our heroine catches her plane,} \\ 0 & \text{otherwise.} \end{cases}$$

$$\textsc{crash} = \begin{cases} 1 & \text{if there is a crash,} \\ 0 & \text{otherwise.} \end{cases}$$

$$\textsc{location} = \begin{cases} 0 & \text{if our heroine ends up stuck at the side of the road,} \\ 1 & \text{if our heroine ends up in Paris,} \\ 2 & \text{if our heroine ends up dead.} \end{cases}$$

CAR and CRASH are what Pearl calls *exogenous* variables; their values are determined by factors outside the model. CATCH and LOCATION are *endogenous* variables; their values are determined by the values of other variables in the model.

For each of the endogenous variables, the model specifies a structural equation. In the plane example, the structural equations are as follows.

$$\textsc{catch} = \textsc{car}$$

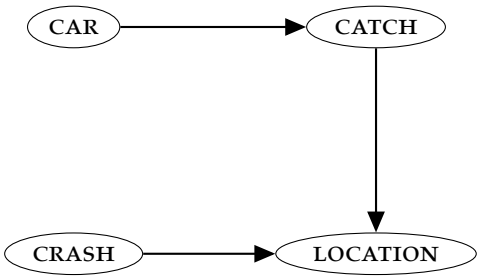
$$\textsc{location} = \begin{cases} 0 & \text{if } \textsc{catch} = 0, \\ 1 & \text{if } \textsc{catch} = 1 \text{ and } \textsc{crash} = 0, \\ 2 & \text{if } \textsc{catch} = 1 \text{ and } \textsc{crash} = 1. \end{cases}$$

(NB: the structural equations are asymmetric. The variable on the left-hand side has its value causally determined by the variables on the right-hand side.)

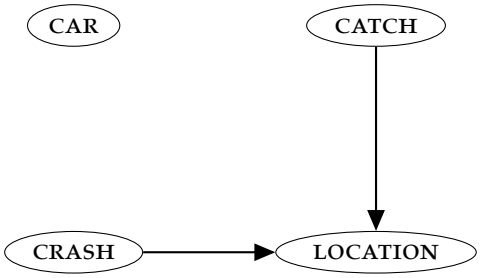
In the plane example, the variables take on the following values.

$$\begin{aligned} \textsc{car} &= 0, \\ \textsc{catch} &= 0, \\ \textsc{crash} &= 1, \\ \textsc{location} &= 0. \end{aligned}$$

We can summarize information about the variables and structural equations using the causal graph in Figure 3a. An arrow from one variable to



(a) The actual model



(b) The submodel generated by intervening to make the antecedent true

Figure 3: Causal graph used to evaluate the counterfactual LETHAL: 'If the heroine had caught that plane, she would be dead by now'.

another indicates that the first variable exerts direct causal influence on the second, but unlike the structural equations, the causal graph doesn't specify the nature of that influence.

Given a pair consisting of a model and an assignment of values to variables in the model, we can use a selection function to assign truth values to conditionals. (This time, the selection function takes in a model, and returns a singleton containing one new model.) Pearl's account is restricted to counterfactuals whose antecedents are either 'literals', which say that a particular variable takes on a particular value, or conjunctions of literals. (So 'the heroine's car breaks down and the plane crashes' is an acceptable antecedent, while 'the heroine's car breaks down or the plane crashes' is not.)

Where $\langle M, V\rangle$ is a model paired with an assignment of values to variables, and A is an antecedent with the appropriate form, we can generate a *submodel* $\langle M_A, V_A\rangle$ by 'intervening' on $\langle M, V\rangle$ to make A true. Intuitively, we can imagine an intervention as an action by someone outside the model who 'reaches in' to make the antecedent true, without tinkering with variables that are causally independent of the antecedent. For instance, a cabbie could intervene to set CATCH $= 1$ by driving our heroine to the airport regardless of whether or not her car has broken down.

Formally, the submodel M_A is a model with the same variables as M, but different structural equations. If X is one of the variables mentioned in

A, and X is endogenous, we delete the structural equation corresponding to X, and make X exogenous instead. (This corresponds to the idea that an intervention makes A true regardless of whether its typical causes obtain; the intervening cabbie enables the heroine to get to the airport whether or not her car is in working order.) We then set the value V_A of each X mentioned in A to the value specified by A. (This corresponds to the idea that the intervention makes the antecedent true.) If a variable is not causally influenced (either directly or indirectly) by any of the variables mentioned in the antecedent, then V_A assigns it the same value as V. (This corresponds to the idea that an intervention is *minimal*, so that only the variables mentioned in the antecedent are directly affected.) Finally, if a variable is causally influenced by one of the variables mentioned in the antecedent, then its value V_A is fixed by the structural equations. (This corresponds to the idea that an intervention is minimal in another sense: it does not interfere with the downstream effects of the variables mentioned in the antecedent.)

We are now ready to evaluate the counterfactual

LETHAL. If the heroine had caught that plane, she would be dead by now.

in our original model. To check whether LETHAL is true in the original model, we intervene to make its antecedent true—i.e., to set CATCH $= 1$. We then check whether the consequent is true (i.e., LOCATION $= 2$) in the resulting submodel.

First, we delete the structural equation for CATCH, turning CATCH into an exogenous variable. Our only remaining structural equation is

$$\text{LOCATION} = \begin{cases} 0 & \text{if CATCH} = 0, \\ 1 & \text{if CATCH} = 1 \text{ and CRASH} = 0, \\ 2 & \text{if CATCH} = 1 \text{ and CRASH} = 1. \end{cases}$$

(The graph for the resulting submodel is shown in Figure 3b.)

Second, we set the values of the variables. The antecedent requires that

$$\text{CATCH} = 1.$$

Since neither CAR nor CRASH is downstream from CATCH, we have

$$\text{CAR} = 0,$$
$$\text{CRASH} = 1.$$

Finally, the value of LOCATION is fixed by the structural equation. Since CATCH $= 1$ and CRASH $= 1$, we have

$$\text{LOCATION} = 2.$$

Therefore, in the submodel, the protagonist is dead, so in the original model, had she caught her plane, she would have been dead.

The procedure described is a type of selection semantics: given an antecedent and a model-valuation pair, we call on a 'submodel' selection function that returns the singleton set of another model-valuation pair (a submodel). Galles and Pearl (1998) argue that this selection semantics is formally equivalent to the closest-worlds account. However, there is a key difference between the two accounts: the selection semantics lets us assigns truth conditions to counterfactuals built up from arbitrary sentences, while the causal modeling account only lets us assign truth values to counterfactuals whose antecedents are literals, or conjunctions of literals. Schulz (2011) and Briggs (2012) propose ways of extending the language to counterfactuals with logically complex antecedents; their proposed theories are logically inequivalent to the closest-worlds semantics. Huber (2013) proposes an alternative way of extending the language that makes it logically equivalent to the closest-worlds account.

5 COUNTERPOSSIBLE CONDITIONALS

Selection semantics has trouble with *counterpossible* conditionals—that is, conditionals whose antecedents are impossible. It counts all counterpossible conditionals as trivially true. Where A is impossible, there are no possible A-worlds. Therefore, if we feed the selection function an impossibility A and a world w and ask it to return a set of possible A-worlds, it returns the empty set. Trivially, all the A-worlds in the empty set are C-worlds, so that trivially $A \mathrel{\Box\!\!\rightarrow} C$ is true in the original world.

But counterpossibles seem to have non-trivial truth conditions: some are true, while others are false. Examples of true counterpossibles include:

> If Hobbes had (secretly) squared the circle, sick children in the mountains of South America at the time would not have cared (Nolan, 1997, p. 544).
>
> If I were a horse, then I would have hooves (Krakauer, 2012, p. 10).
>
> If wishes were horses, beggars would ride (Krakauer, 2012, p. 10).
>
> If intuitionistic logic were the correct logic, then the law of excluded middle would no longer be unrestrictedly valid (adapted from Brogaard & Salerno, 2013).

Corresponding examples of false counterpossibles include:

> If Hobbes had (secretly) squared the circle, sick children in the mountains of South America at the time would have taken notice.

If I were a horse, then I would have scales.

If wishes were horses, no one would own any horses.

If intuitionistic logic were the correct logic, then the law of excluded middle would still be unrestrictedly valid.

Assigning non-trivial truth values to counterpossibles doesn't just capture linguistic intuitions; it also enables counterpossibles to do valuable philosophical work. Non-trivial counterpossibles help us assess rival philosophical, mathematical, and logical theories by telling us what would follow if those theories were true (Krakauer, 2012; Brogaard & Salerno, 2013; Nolan, 1997). They explain how necessary events and omissions of impossible events are causally relevant to the actual world—how a mathematician's failure to disprove Fermat's Last Theorem prevented her from getting tenure, how my failure to be in two places at once caused me to miss a colloquium talk, or how the copresence of a mental property and its subvening physical property can result in a subject's raising his arm (Bernstein, 2016). They can be used to give an account of essences: an essential property is one such that, if the bearer had lacked it, then the bearer would not have existed (Brogaard & Salerno, 2013, 2007). (Non-trivial counterpossibles save this account from certain implausible commitments—e.g., that living in a world where $2 + 2 = 4$ is trivially part of everyone's essence.)

Not everybody agrees that counterpossibles have non-trivial truth values, however. Williamson (2007, p. 172) argues that apparent examples of non-trivial counterpossibles collapse under closer scrutiny. In a slight variant on Williamson's example,[8] imagine that a student is mulling over a graded arithmetic test. Of the 12 problems on the test, the student has gotten the last one wrong: 'what is $5 + 7$?' The student, who answered '11', laments: 'If only $5 + 7$ were 11, I would have gotten a perfect score!' This seems to be true, and furthermore, it seems false that if $5 + 7$ were 11, the student would have gotten one of the problems wrong. But appearances are deceptive. Suppose that $5 + 7$ were 11. Then in answering all the problems right, the student would have given five right answers followed by seven more right answers, for a total of 11 right answers. Since there are 12 problems on the test, the student would have gotten one problem wrong after all. (For a rebuttal of Williamson's argument, see Salerno and Brogaard, 2007.)

5.1 *Impossible Worlds*

Nolan (1997) gives an account of counterpossibles by supplementing the closest-worlds account with impossible worlds—ways the world couldn't

8 Thanks to Sharon Berry for suggesting this version in conversation.

be. We can then say that $A \boxrightarrow B$ is true at w just in case B is true at all the closest possible or impossible A-worlds to w. Two questions then arise: what are impossible worlds, and what makes them closer to or further away from the actual world?

The ontology of impossible worlds has spawned its own literature: they may be collections of individuals like our actual world (Yagisawa, 2010), or they may be sets of sentences in some suitable language (Hintikka, 1975; Melia, 2001; Sider, 2002; see Berto, 2013, for a general overview and discussion.) Another pressing question for theorists of counterpossibles concerns the logical structure of impossible worlds. Is it the case that for every set of sentences, there is some impossible world where all and only the sentences in the set are true? Or is there more logical structure we can impose on impossible worlds?

Proponents of impossible worlds typically don't require that the impossible worlds be closed under classical logical consequence—in other words, they don't require that whenever some propositions are true at an impossible world, all the classical logical consequences of those propositions are true at the world too. If impossible worlds had to be closed under classical logical consequence, then whenever A was impossible by the rules of classical logic, $A \boxrightarrow C$ would be trivially true. Nolan (1997, p. 547) argues that we should not require impossible worlds to be closed under any kind of logical consequence, since for every putative logical truth, there are non-trivial facts about what the world would be like if that logical truth did not obtain. A similar line of reasoning suggests that some impossible worlds have truth-value gluts: we can speculate about what would happen if there were true contradictions, so there must be impossible worlds at which there are true contradictions.

Bjerring (2013) argues that some impossible worlds have truth-value gaps. Otherwise, he argues, our theory of counterpossibles would misclassify certain conditionals as true, such as this one: 'If intuitionistic logic were correct, then the Law of Excluded Middle would hold.' (The Law of Excluded Middle says of every proposition that either it or its negation holds; intuitionists famously deny it.)

What about closeness? Nolan (1997) proposes the

> Strangeness of Impossibility Condition. Any possible world is more similar [closer] to the actual world than any impossible world (Nolan, 1997, p. 550).

The Strangeness of Impossibility Condition ensures that where A is a possible proposition, supplementing the closest-worlds account with impossible worlds has no effect on how we evaluate $A \boxrightarrow C$. So long as A is possible, the set of closest possible A-worlds coincides with the set of closest possible or impossible A-worlds.

Bjerring (2013, p. 348) proposes another constraint on closeness, which implicitly relativizes closeness to the antecedent of a counterfactual. Given a collection of logical systems $L_1, L_2, \ldots, L_n$, where L_1 is classical logic, and where W_{L_i} is the set of worlds deductively closed under L_i's entailment relation, Bjerring endorses the

RELATIVE CLOSENESS CONDITION. For any counterfactual whose antecedent presupposes that some logic L_i is correct (true, adequate), a world in modal space W_{L_i} is closer to the actual world than any world in modal space W_{L_j} , where $W_{L_i} \neq W_{L_j}$, and where $i \geq 1$ and $j > 1$.[9]

Brogaard and Salerno (2013) develop a theory on which impossible worlds are close to the actual world to the extent that they

1. minimize discrepancies with relevant background facts about the actual world (where the relevance of background facts is fixed by context), and

2. minimize violations of relevant *a priori* entailment (where relevant *a priori* entailment is spelled out in more detail in the paper).

As an illustration of these conditions, Brogaard and Salerno use them to evaluate the counterpossible conditional 'if water had not been H_2O, then water would have been a monkey'. This counterpossible is false. Their theory delivers the correct verdict, they claim, because it is *a priori* that water is not a monkey.

To derive this verdict, they consider two impossible worlds where the antecedent is true. At w_1, water is some chemical compound XYZ (different from H_2O), while at w_2, water is a monkey.

w_1	w_2
water is not H_2O	water is not H_2O
water is XYZ	water is a monkey

Since there are more *a priori* truths that hold at w_1 than at w_2, and since both agree with the actual world about the same number of propositions, w_1 is closer to the actual world than w_2. (Brogaard and Salerno tacitly assume that there are no antecedent worlds closer to the actual world than

9 As stated by Bjerring, the Relative Closeness Condition seems to presuppose that W_{L_i} and W_{L_j} do not intersect. We can get rid of this presupposition by modifying the condition slightly:

RELATIVE CLOSENESS CONDITION*. For any counterfactual whose antecedent presupposes that some logic L_i is correct (true, adequate), a world in modal space W_{L_i} is closer to the actual world than any world outside W_{L_i}.

w_1 or w_2.) Thus, at least one of the closest impossible worlds where water is not H_2O is one where water fails to be a monkey, so the conditional is false at the actual world.

5.2 *Relevant Logic*

Relevant logics are motivated by the thought that the conditional 'if A, then C' claims that the truth of A is connected to the truth of C. Relevant logics originated as rivals to the material conditional account, on which the conditional 'if A, then C' is true just in case A is false or C is true (see Section 6). However, some of the same intuitions that favor relevant logics over the material conditional account also favor them over the closest-worlds account. After all, the reason it seems wrong to say 'if Hobbes had squared the circle, sick children in the mountains of South America would have cared' is that there is no connection between Hobbes's squaring the circle and the interests of sick South American children. Likewise, the reason it seems right to say 'if I were a horse, I would have hooves' is because something's being a horse is connected to its having hooves.

Relevant logics are often characterized in proof-theoretic terms. But Routley and Meyer (1973, 1972a, 1972b) develop a versatile semantics for the conditionals of relevant logics, which generalizes the strict conditional semantics of Section 4.1. Recall that on the strict conditional interpretation, $A \boxdot\!\!\rightarrow C$ is true at w just in case C is true at all possible A-worlds (relative to w). We can rewrite the selection function in terms of a two-place accessibility relation among worlds: we say that Rwx just in case world x is possible according to world w, and that $f(A, w)$ is the set of all A-worlds x such that Rwx.

Routley and Meyer interpret the conditional in terms of a three-place accessibility relation among worlds. 'If A, then C' is true at w just in case, for all worlds x and y such that $Rwxy$ and x is an A world, C is true at y. Different restrictions on relation R generate different relevant logics. (For some logics, we need impossible worlds where both a sentence and its negation fail to be true, or impossible worlds where both sentence and its negation are true.)

This three-place R relation is formally useful, but does it mean anything? Beall et al. (2012) propose three interpretations of $Rwxy$, which spring from different ways of grouping w, x, and y.[10] All three interpretations can be illustrated with the conditional

THERMITE. If you light a bucket of thermite with a titanium fuse, then a huge explosion will ensue.

10 For a discussion of other ways of interpreting the ternary relation, with references, see Jago (2013).

GROUPING THE SECOND AND THIRD WORLDS TOGETHER: $Rw\langle xy\rangle$. 'If A, then C' says at the actual world w, there are no counterexamples where A is true and C is false. We typically think of counterexamples as involving a single world which makes some things true and other things false, but relevant logicians split the labor between two worlds x and y, so that whatever holds at x is true, while whatever fails to hold at y is false. In the example of THERMITE, we might think of potential counterexamples as divided into an earlier part x, when a bucket of thermite may or may not be lit with a titanium fuse, and a later part y, when there may or may not be an explosion. If the actual world w admits some possible two-part scenarios that begin with the lighting of thermite with a titanium fuse, but fails to end in a huge explosion, then these scenarios are counterexamples that falsify THERMITE.

GROUPING THE FIRST AND SECOND WORLDS TOGETHER: $R\langle wx\rangle y$. 'If A, then C' says that using one's current information to draw inferences from A will yield the information that C. To say that $Rwxy$ is to say that when the rules of w are applied to the information in x, it is possible to infer y (or some information that entails y). In the case of THERMITE, we can imagine w as a parcel of information specifying the actual laws of nature, and x as another parcel of information specifying that a bucket of thermite has been lit with a titanium fuse. If sticking these parcels of information together licenses the conclusion that there has been a huge explosion (and does so no matter how we fill in x, the information that the thermite has been lit), then the conditional THERMITE is true.

GROUPING THE FIRST AND THIRD WORLDS TOGETHER: $Rw\rangle x\langle y$. 'If A, then C' says that C is necessary relative to A, or that C is necessary in an A-ish way. The conditional THERMITE does not say it is absolutely necessary that a huge explosion will ensue. The world w may permit a possible scenario y in which no huge explosions occur. However, once we enrich w with some additional information x, specifying that a bucket of thermite has been lit with a titanium fuse, we can consider what is possible under that supposition. If there is some way of filling in the antecedent that makes y a possibility, then y is possible not just absolutely, but under the supposition that the antecedent of THERMITE is true.

Mares and Fuhrmann (1995) propose a theory of counterfactuals that combines the closest-worlds interpretation of the selection function with the relevant interpretation of the conditional: $A \boxdot\!\!\rightarrow B$ is true at a world w just in case the relevant conditional 'if A, then B' is true at all closest A-worlds to w. Mares (1994) argues that this theory has useful applications to conditional analyses of causation, and to theories of conditional obligation.

6 THE MATERIAL CONDITIONAL ACCOUNT OF INDICATIVES

According to the material conditional account defended by Grice (1989) and Jackson (1987), an indicative conditional $A \to C$ is true just in case either its antecedent A is true, or its consequent C is false. (The material conditional account is almost always offered as a theory of indicative conditionals alone, since counterfactual conditionals with false antecedents can be false. Even though I don't keep a horse, it is false that if I were to keep a horse, it would breathe fire.) The material conditional account has a simple explanation for the apparent validity of all the the inferences discussed in Section 2 (modus ponens, modus tollens, conditional proof, strengthening the antecedent, transitivity, contraposition, and simplification): these inferences really are valid.

Furthermore, there are persuasive arguments for the conclusion that an indicative conditional $A \to C$ is true if and only if the corresponding material conditional 'not A or C' is true. Suppose the indicative conditional is true. Then it can't have a true antecedent and a false consequent; that would be a violation of modus ponens. So the indicative conditional entails the material conditional. But when I know that either C holds or A doesn't, I can infer that if A, then C. So the material conditional entails the indicative. (Stalnaker, 1975, p. 136, calls this the direct argument.) Since the material and indicative conditionals entail each other, they must be equivalent.

Gibbard (1981) provides a formal argument for the equivalence of the indicative and material conditionals based on three logical principles. Where 'not A' is abbreviated $\neg A$ and 'A or B' is abbreviated $A \vee B$, the principles are:

PSEUDO MODUS PONENS. $A \to C$ entails $\neg A \vee C$.

IMPORT-EXPORT. $A \to (B \to C)$ is equivalent to $(A \wedge B) \to C$.

CONDITIONAL PROOF. If A entails C, then $A \to C$ is a logical truth.

To show that $A \to C$ and $\neg A \vee C$ are equivalent, Gibbard only needs to show that each entails the other. By Pseudo Modus Ponens, $A \to C$ entails $\neg A \vee C$. The proof that $\neg A \vee C$ entails $A \to C$ is as follows.

1. $((\neg A \vee C) \wedge A)$ entails C. (By tautological reasoning.)
2. It is a truth of logic that $((\neg A \vee C) \wedge A) \to C$. (By 1 and Conditional Proof.)
3. It is a truth of logic that $(\neg A \vee C) \to (A \to C)$. (By 2 and Import-Export.)

4. It is a truth of logic that $\neg(\neg A \vee C) \vee (A \rightarrow C)$. (By 3 and Pseudo Modus Ponens.)

5. $(\neg A \vee C)$ entails $(A \rightarrow C)$. (By 4 and tautological reasoning.)

Despite these points in its favor, the material conditional account faces substantial difficulties. It seems to yield wrong predictions about logical validity, often called 'paradoxes of material implication'.

For example,the material conditional account entails that all of the following are truths of logic:

> Either the unburied dead will walk the Earth if I bury a chicken head in my backyard, or the unburied dead will walk the Earth if I fail to bury a chicken head in my backyard (McGee, 2005).
>
> Either you are virtuous if you are rich, or you are rich if you are virtuous.
>
> One of these three things holds: if you grant voting rights to children, you will grant them to guinea pigs; if you grant voting rights to guinea pigs, you will grant them to inanimate objects; or if you grant voting rights to inanimate objects, you will take them away from adult human beings.

Furthermore, the material conditional account entails that all of the following inferences are valid. (The proof of God's existence is due to Edgington, 1986.)

1. I will not do my chores today.

$\therefore$ If I do my chores today, then the world will implode.

1. Dinner will be delicious.

$\therefore$ If I burn the veggie burgers and pour sand into the sweet potatoes, then dinner will be delicious.

1. If God does not exist, then it's not the case that if I pray, my prayers will be answered.
2. I do not pray.

$\therefore$ God exists.

In addition to yielding bad predictions about validity, the material conditional account yields bad predictions about the probabilities of conditionals. Suppose I draw a card at random from a 52-card deck. The material conditional 'either I do not draw a red ace, or I draw the ace of hearts' has probability 51/52. (The only way for me to make it false is to draw the ace of diamonds.) Therefore, by the material conditional account, I should assign probability 51/52 to the indicative conditional 'if I draw a

red ace, then it will be the ace of hearts'. But the indicative conditional 'if I draw a red ace, then it will be the ace of hearts' should get probability 1/2, since half the time when I draw a red ace, it will be an ace of hearts.

More generally, the material conditional account falls afoul of

THE THESIS. Whenever A and C are propositions, the probability of the indicative conditional $A \rightarrow C$ is equal to the conditional probability of C given A, understood as

$$Pr(A|C) = \frac{Pr(A \wedge C)}{Pr(C)}.$$

THE THESIS is a plausible way of unpacking the so-called *Ramsey test*, based on a famous remark by Ramsey (1978, 143n):

> If two people are arguing 'If p will q?'; and are both in doubt as to p, they are adding p hypothetically to their stock of knowledge and arguing on that basis about q; so that in a sense 'If p, q' and 'If p, [not q]' are contradictories.

Unfortunately, the material conditional account is straightforwardly incompatible with THE THESIS, and with the Ramsey test more generally. The probability that a material conditional is true is not, in general, the conditional probability of the consequent given the antecedent. (The probability of the material conditional may be anywhere between that conditional probability and 1.) Furthermore, where A is highly unlikely, the material conditional 'not A or C' is both highly believable and highly assertible, whether or not adding A to one's stock of knowledge would justify a high degree of confidence in C.

Grice (1989) and Jackson (1987) explain these wrong predictions by distinguishing between true sentences and sentences that can appropriately be asserted. According to Grice, in a situation where I will not do my chores today, it is technically true that if I do my chores today, then the world will implode. Likewise, in a situation where dinner will be delicious, it is technically true that if I burn the veggie burgers and pour sand into the sweet potatoes, then dinner will be delicious. Nonetheless, it is misleading to assert a conditional when I know that its antecedent is false, or when I know that its consequent is true, because it is misleading to assert a weak claim when I could have asserted a stronger one. Refusing to assert the stronger claim is liable to mislead my audience into thinking that I do not know it. The supposedly paradoxical arguments are valid. When their premises are true, their conclusions may be bad, but this does not make their conclusions false.

Grice's proposed mechanism for explaining away the problem is useful in other domains: it can explain why some non-conditional assertions are

misleading. For instance, if you ask where John is, and I know that he is in the library, it is misleading for me to reply 'He is either at the pub, or in the library.' A similar trick works for negated conjunctions, as an example by D. Lewis (1976) shows. If I point out a harmless mushroom that I plan to keep for myself, and remark 'You won't eat that and live', knowing that my assertion will prevent you from eating it, then I am guilty of misleading you, though what I say is technically true.

Jackson (1987) is not satisfied with Grice's explanation, since sometimes, it is all right to assert an indicative conditional even if you know that the antecedent is false, or the consequent is true. I know that Oswald killed Kennedy, but can nonetheless assert that if Oswald didn't kill Kennedy, someone else did. Jackson has a different explanation for why technically true conditionals might sound wrong. While the material conditional account captures the truth conditions of an indicative conditional, the meaning of 'if... then...' goes beyond its truth conditions. Built into the meaning of an English indicative conditional is the implication that it would still be appropriate to assert the material conditional, even if its antecedent were known. (Jackson calls this feature 'robustness'.) The Oswald-Kennedy conditional is robust, because even if I had reason to doubt that Oswald killed Kennedy, I would still have good reasons to believe that either Oswald or someone else killed him.

7 THE NO TRUTH VALUES (NTV) ACCOUNT OF INDICATIVES

Suppose you are convinced that the material conditional account gives the wrong truth conditions for the indicative conditionals. You might hope that there was some other account of the truth conditions for indicative conditionals—one that could better explain the truth of THE THESIS. Unfortunately, a collection of so-called 'triviality theorems' suggests that no truth conditions whatsoever will do the trick. Triviality theorems motivate Edgington (1986, 1995) and Appiah (1985) to claim that indicative conditionals lack truth values altogether. (Edgington, 2008, goes on to develop a Y-shaped theory on which counterfactual conditionals also lack truth values altogether.)

In general, triviality theorems show that if THE THESIS is true in general, then every probability function is *trivial*: it assigns positive probability to at most two mutually exclusive alternatives. But it is absurd to claim that every probability function is trivial. (Here is a non-trivial probability function: the one that assigns probability 1/6 to each possible outcome of the roll of a single die.) Therefore, we must reject THE THESIS.

To see how triviality theorems work, we can consider an early result by D. Lewis (1976), illustrated by system of diagrams adapted from Edgington (1995). Edgington visualizes probabilities using rectangles, divided into

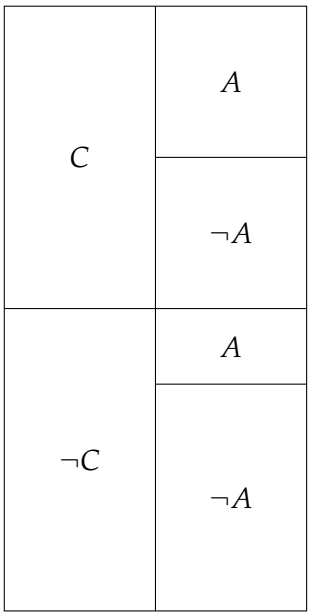


Figure 4: A probability space

horizontal segments representing propositions. The height of a segment represents the probability of the corresponding proposition; the entire rectangle is normalized to have height 1. In Figure 4 the proposition C has probability $1/2$. C is subdivided into the propositions $A \wedge C$ (probability $1/4$) and $A \wedge \neg C$ (probability $1/4$). $\neg C$ (also with probability $1/2$) is subdivided into the propositions $\neg C \wedge A$ (probability $1/8$) and $\neg C \wedge \neg A$ (probability $3/8$).

Figure 5 shows how to calculate the probability of A conditional on C by erasing the bottom half of the diagram, and stretching out the remaining part of the rectangle so its height is 1 (in effect multiplying the height of each of its sub-regions by $\frac{1}{Pr(C)}$). The new height of the A region is $Pr(A|C)$.

According to the Law of Total Probability (illustrated in Figure 6), for any two propositions X and Y,

$$Pr(Y) = Pr(Y|X) \times Pr(X) + Pr(Y|\neg X) \times Pr(\neg X). \quad (1)$$

Consider any two propositions A and C such that $P(A \wedge C) > 0$, and $P(A \wedge \neg C) > 0$. Plugging in A for X and $A \rightarrow C$ for Y in Equation 1 yields:

$$Pr(A \rightarrow C) = Pr(A \rightarrow C|C) \times Pr(C) + Pr(A \rightarrow C|\neg C) \times Pr(\neg C). \quad (2)$$

In other words, we can split the probability space into a C part and a $\neg C$ part, and figure out the probability of $A \rightarrow C$ by averaging its

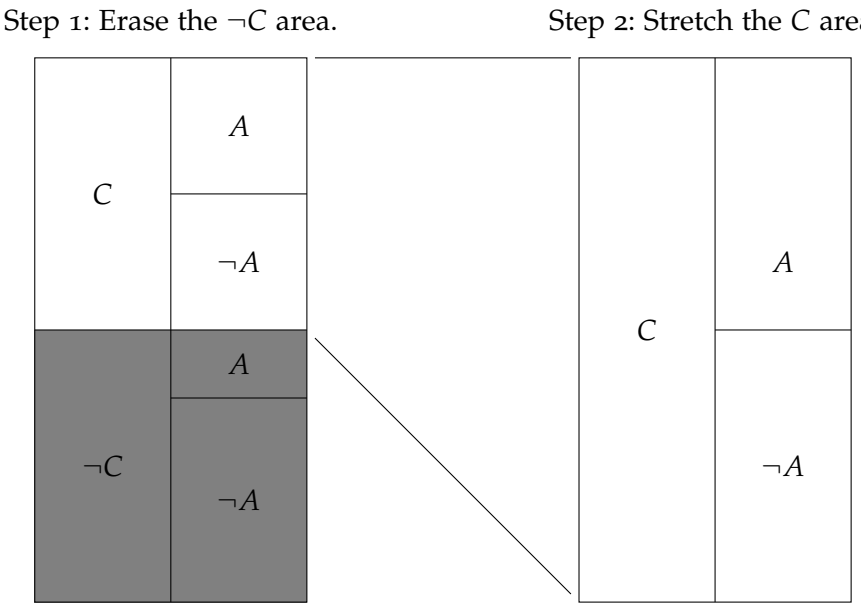


Figure 5: Calculating conditional probability

probabilities conditional on each part, a procedure illustrated in Figure 7. Consider the probability distribution Pr_C such that for all propositions X, $Pr_C(X) = Pr(X|C)$ (shown in the top center of Figure 7). Using the fact that $Pr_C(A) > 0$, the fact that $Pr_C(C) = 1$, and the definition of conditional probability, we can show that

$$Pr_C(C|A) = 1. \tag{3}$$

Thus, by THE THESIS and Equation 3,

$$Pr_C(A \rightarrow C) = 1. \tag{4}$$

By the definition of Pr_C and Equation 4,

$$Pr(A \rightarrow C|C) = 1. \tag{5}$$

Likewise, when we consider the probability distribution $Pr_{\neg C}$ such that for all X, $Pr_{\neg C}(X) = Pr(X|\neg C)$ (shown in the bottom center of Figure 7), we see by the fact that $Pr_C(A) > 0$, the fact that $Pr_C(C) = 1$, and the definition of conditional probability that

$$Pr_{\neg C}(C|A) = 0. \tag{6}$$

Thus, by THE THESIS, and Equation 6,

$$Pr_{\neg C}(A \rightarrow C) = 0. \tag{7}$$

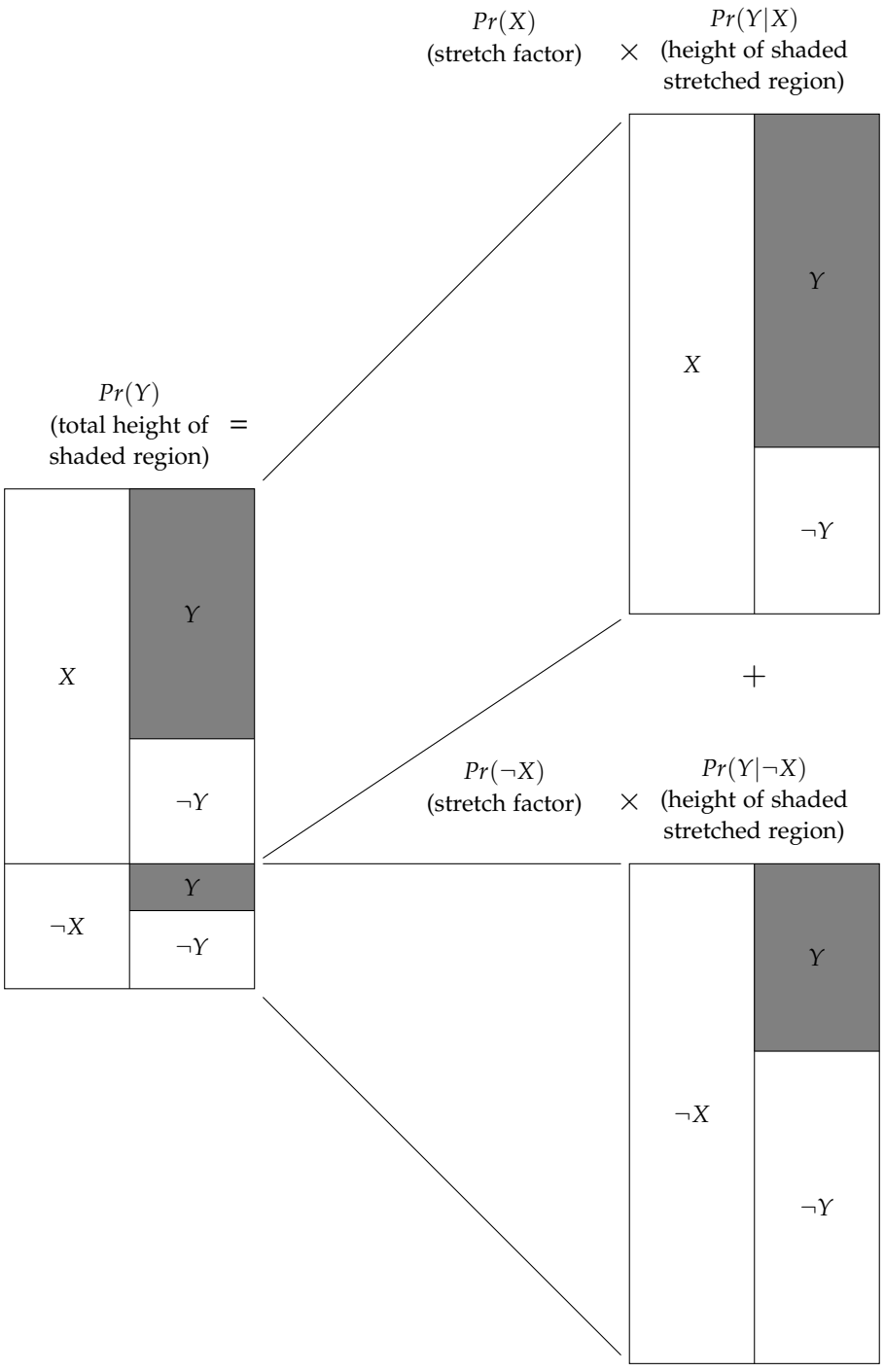


Figure 6: The Law of Total Probability

And by the definition of $Pr_{\neg C}$ and Equation 7,

$$Pr(A \rightarrow C|\neg C) = 0. \tag{8}$$

Using Equation 5 and Equation 8 to make the appropriate substitutions into equation Equation 2, we get:

$$Pr(A \rightarrow C) = 1 \times Pr(C) + 0 \times Pr(\neg C) = Pr(C). \tag{9}$$

But by THE THESIS,

$$Pr(A \rightarrow C) = Pr(C|A). \tag{10}$$

Substituting $Pr(C|A)$ for $Pr(A \rightarrow C)$ on the left-hand side of Equation 9, we get:

$$Pr(C|A) = Pr(C) \tag{11}$$

—in other words, A and C are probabilistically independent.

The above proof shows that Equation 11 holds for arbitrary propositions A and C, provided both $Pr(A \wedge C)$ and $Pr(A \wedge \neg C)$ are both greater than 0. Therefore Equation 11 should hold for all pairs of propositions A and C such that $Pr(A \wedge C)$ and $Pr(A \wedge \neg C)$ are both greater than 0. But this is only possible in trivial probability spaces. So one of our assumptions must have gone wrong, and the natural place to pin the blame is on THE THESIS.

There are various possible ways out of Lewis's triviality theorem. The proof assumes that the conditional $A \rightarrow C$ has a single set of truth conditions, which remain stable across Pr, Pr_C, and $Pr_{\neg C}$. Defenders of THE THESIS might reject this assumption and claim that the truth-conditions of conditionals are context-dependent. The proof also assumes that THE THESIS holds for all probability functions and all conditionals. Defenders of THE THESIS might retreat and claim that it is true for only some conditionals, or some probability functions.

Unfortunately, both escape routes are treacherous. New triviality theorems can be derived from much weaker assumptions; for a helpful survey, see Hall and Hájek (1994). There are even triviality results that use non-probabilistic variants of THE THESIS (Gärdenfors, 1988), and trivializing versions of THE THESIS that apply to counterfactuals rather than indicatives (Williams, 2012). On a slightly more optimistic note, *non*-triviality results can be obtained by adopting (sufficiently weak) non-classical logics (Morgan & Mares, 1995).

Another way out of Lewis's triviality theorem is to reject THE THESIS. Kaufmann (2004) produces examples of indicative conditionals in English that intuitively seem to violate THE THESIS, and Douven and Verbrugge (2013) provide experimental evidence that English speakers' judgments about indicative conditionals violate THE THESIS.

If probability is probability of truth, defenders of the NTV view should reject THE THESIS too. However, defenders of the NTV view typically defend

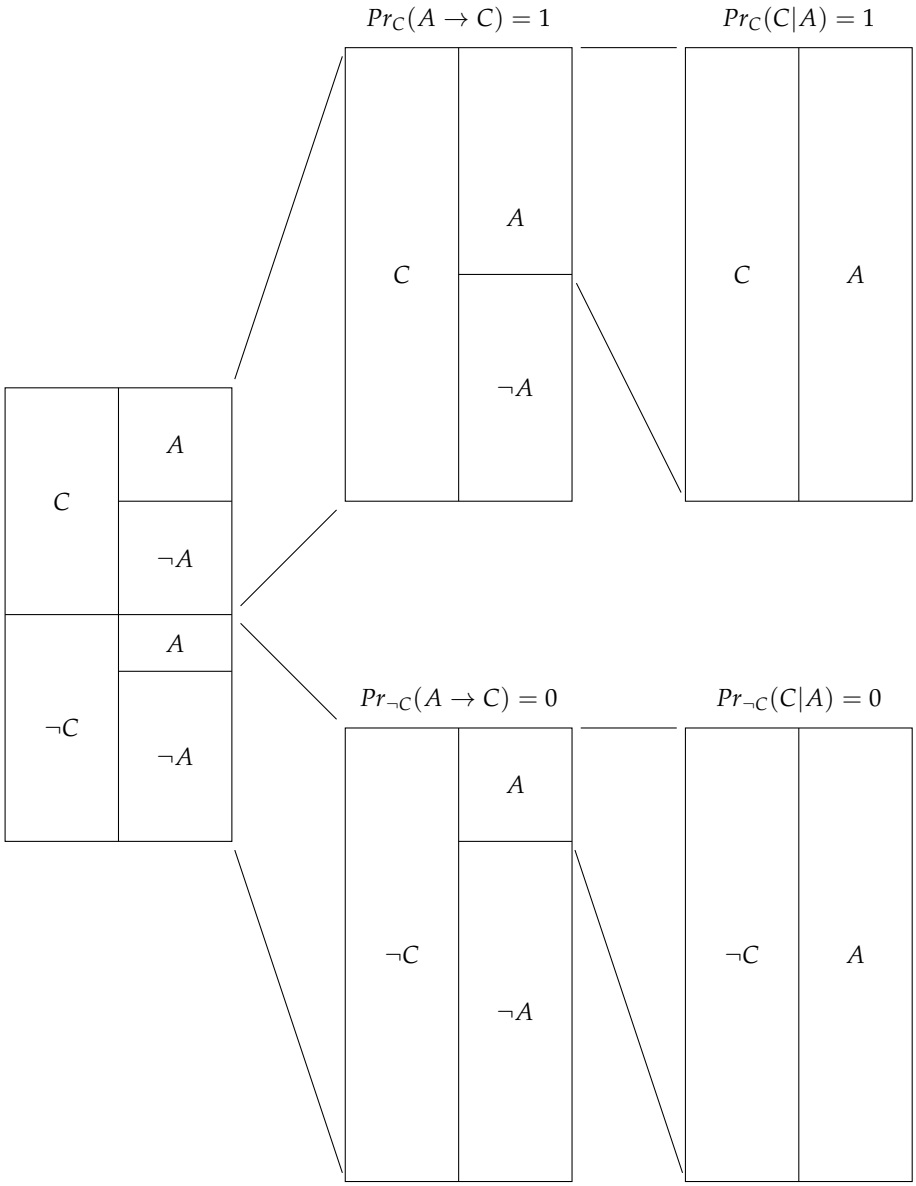


Figure 7: The Lewis triviality theorem illustrated

versions THE THESIS, but adopt alternative interpretations of 'probability', on which the probability of a conditional is not the probability of its truth.

Calling on alternative theories of probability makes sense: probability is a versatile explanatory tool, and the NTV theory has plenty of explaining to do. In particular, the NTV theory needs to explain why conditionals seem to have the features of truth-evaluable statements. It is sometimes reasonable to believe a conditional—but ordinarily, to believe something is to believe that it is true. Likewise, it is sometimes reasonable to assert a conditional—but ordinarily, to assert something is to claim that it is true. Arguments with conditionals in their premises and conclusions are sometimes valid and sometimes invalid—but ordinarily, a valid argument is one that cannot have true premises and a false conclusion, and it's not clear how to fruitfully apply the concept of validity when a premise or conclusion lacks truth conditions altogether.

Adams (1975) and Edgington (1986) give a probabilistic account of belief in conditionals. Belief comes in degrees, which are measured by probabilities. A person's degree of belief in a conditional is simply her conditional degree of belief in its consequent on its antecedent.

Adams (1975) gives a probabilistic account of validity for conditionals. An argument is said to be probabilistically valid just in case it is impossible for its premises to be probable and its conclusion improbable. More precisely, an argument from premises $P_1, P_2, \ldots, P_n$ to conclusion C is valid just in case, for every real number $\epsilon > 0$, there is a real number $\delta > 0$ such that, if each of $P_1, P_2, \ldots, P_n$ has probability greater than $1 - \delta$, then C has probability at least $1 - \epsilon$.

Adams' definition of validity coincides with the classical definition where $P_1, P_2, \ldots, P_n$ and C are conditional-free sentences, and lets us define validity for arguments containing simple conditionals. The theory is built to handle only simple conditionals, and does not let us assess validity for arguments containing compound sentences with conditionals as parts. McGee (1989) extends Adams' theory to cover compounds of conditionals.

Edgington (1995) gives a non-probabilistic account of what it is to assert a conditional: it is to assert the consequent if the antecedent is true, and to assert nothing otherwise. She argues that her account assimilates conditional assertions to a larger class of conditional speech acts, including:

CONDITIONAL QUESTIONS. 'If he phones, what shall I say?'

CONDITIONAL COMMANDS. 'If he phones, hang up.'

CONDITIONAL PROMISES. 'If he phones, I promise not to be rude.'

CONDITIONAL AGREEMENTS. 'If he phones, we're on for Sunday.'

CONDITIONAL OFFERS. 'If you phone, you can have a 20% discount.'

Any speech act whatsoever, she claims, can be performed conditionally or unconditionally. We can think of conditionals as 'speech act bombs' primed to detonate when and only when the antecedent is true (see Egan, 2009).

8 DYNAMIC SEMANTICS

So far, we've seen several accounts of conditionals that posit more to their meanings than truth conditions—either because conditionals have no truth conditions (on the NTV account) or because their truth conditions are not sufficient to determine when they can reasonably be asserted (on the material conditional account). Enter dynamic semantics, which provides new tools for modeling meaning.

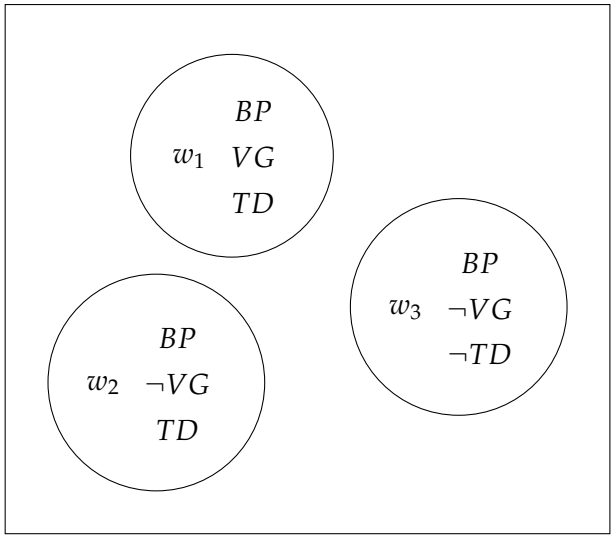
Dynamic semantics explains the meanings of sentences by appeal to a *conversational context*—a set of background assumptions taken for granted by all the participants in a conversation. For instance, if a group of friends is discussing where to go for lunch, the conversational context might include the information that among the nearby restaurants are Veggie Garden and Buddha's Palace. The *context set* is a set of worlds compatible with those background assumptions (see Stalnaker, 1999, p. 84).[11]

The conversational context changes as the conversation progresses, and the context set shrinks and grows accordingly. When a participant makes an assertion, then the content of the assertion is added to the context, and all the worlds incompatible with what is asserted are eliminated from the context set. For instance, if someone asserts that Veggie Garden is open, then the worlds where Veggie Garden is closed are eliminated from the context set.

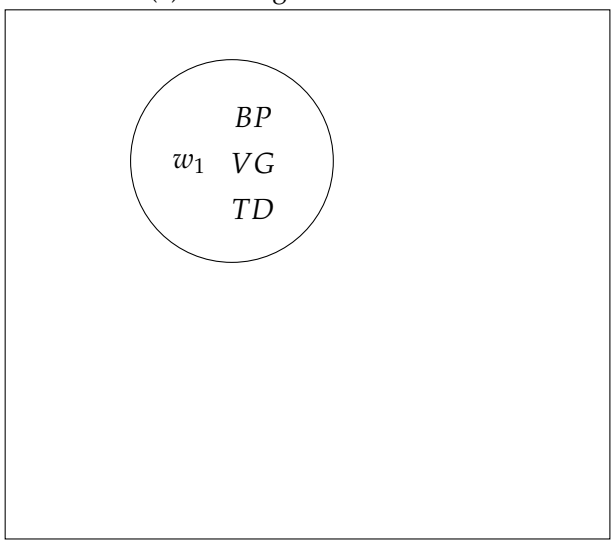
Figure 8 depicts the effect of asserting 'Veggie Garden is open today' on the context set. The original context set is shown in the rectangle at the top of the figure: the circles depict worlds. Each world is labeled with a set of propositions true at that world: 'BP' stands for 'Buddha's Palace is open'; 'VG' stands for 'Veggie Garden is open'; and 'TD' stands for 'we can get gluten-free tofu dogs'.

Notice that some assertions have no effect on the context set. If someone were to assert 'Buddha's Palace is open', none of the worlds in the context set would be eliminated. This is because 'Buddha's Palace is open' is already acceptable in the original context—it follows from what is accepted.

11 To give a complete theory of conditionals, the conversational context will need to include more information than just the context set. Other proposed parameters include a probability function or set of probability functions (Yalcin, 2007, 2012b), and a function that ranks worlds from most to least likely (Spohn, 2015). However, I focus my exposition on the context set to provide a simple illustration of the main ideas.



(a) The original context set



(b) The context set after an assertion of 'Veggie Garden is open'

Figure 8: The effects of an assertion on the context set

Starr (2014) proposes that within this dynamic semantics framework, conditionals can be understood as *tests*, along the lines of the Ramsey test. To determine the effect of asserting a conditional 'if A, then B' on a context c, we first suppose A, by considering the context $c[A]$ that results from adding A to c. We then check whether B is true under the supposition. If B is true at $c[A]$, then c 'passes' the test, and c remains unchanged. Otherwise, c 'fails' the test. If a conditional passes the test, it is acceptable in the original context.

This characterization of acceptability, by itself, is not enough to determine the effect of uttering a conditional in a context where it is not already acceptable. For instance, suppose you go to pet a dog, and I say 'if you pet it, it will bite.' This conditional doesn't follow from our shared background information, but you can use it to rule out possibilities—in particular, those possibilities where you pet the dog and it does not bite. What explains the relationship between my utterance and the corresponding change to the context set? In very broad terms, uttering a conditional should change the context set so that the conditional becomes acceptable, and the change involved should be the smallest one that does the job. There are multiple ways of spelling out what constitutes a minimal change to contextual information, but the part of the account that deals with acceptability can be separated from the part that deals with context change.

To illustrate the concept of a conditional test, consider the conditional 'if Veggie Garden is open, then we can get gluten-free tofu dogs', as asserted in the context depicted by Figure 8a. To perform the test, we first create a new context, by augmenting the old context with the information that Veggie Garden is open; the resulting context set is depicted in Figure 8b. We then check whether, in the new context, 'we can get gluten-free tofu dogs' is acceptable. If so, the old context passes the test, and the conditional is acceptable in the old context; otherwise, the old context fails the test, and the conditional is not acceptable in the old context. Starr extends his account to handle counterfactuals (which use a modified test in which the context set is expanded with extra possibilities before adding the information in the antecedent).

Other theorists offer context-dependent truth conditions for conditionals using the tools of dynamic semantics. Stalnaker (1975) and Williams (2008) defend a modified closest-worlds theory of indicative conditionals, where, if w is a world in the context set, every world in the context set is stipulated to be closer to w than every world outside it. Gillies (2007) and von Fintel (2001) propose strict conditional theories of counterfactuals, where a set of salient worlds is fixed by the context. New worlds are added to the set as the conversation goes on; in particular, if someone asserts a conditional whose antecedent is false at all the salient worlds, the set is expanded to include at least one world compatible with the antecedent.

9 CONDITIONALS AS MODAL RESTRICTORS

According to Kratzer (2012, p. 86), many of the above views of conditionals are 'based on a momentous syntactic mistake.' Contrary to popular opinion, she claims, 'There is no two-place *if…then* connective in the logical forms for natural languages.' Instead, conditionals restrict modal operators.

One can think of modal operators as quantifiers over possible worlds: to say that necessarily $2+2=4$ is to say that in all possible worlds, $2+2=4$; to say that possibly pigs fly is to say that in some possible world, pigs fly; and to say that it will probably rain is to say that in most possible worlds (on some suitable way of measuring 'most'), it rains. Like quantifiers, modal operators can be restricted. To say that necessarily, if the Peano axioms are true, then $2+2=4$, is to say that in all possible worlds where the Peano axioms are true, $2+2=4$. Likewise, to say that if pigs had hollow bones, then possibly pigs would fly, is to say that in some possible world where pigs have hollow bones, pigs fly, and to say that if there are cumulus clouds on the horizon, it will probably rain, is to say that in most possible worlds where there are cumulus clouds on the horizon, it will rain.

The modal restrictor view is a generalization of work by D. Lewis (1975) who notes that conditionals can be used to restrict quantifiers. Consider the following class of examples.

Sometimes Always Usually Never	if a farmer owns a donkey, she feeds it carrots.

The quantifiers 'sometimes', 'always', 'usually', and 'never' are what Lewis calls *unselective quantifiers*. To say that always, farmers feed donkeys carrots is to say that for all ways of assigning a farmer to x and a donkey to y, x feeds y carrots. To add the clause 'if a farmer owns a donkey' is to restrict the quantifier, so that it ranges only over cases where farmer x owns donkey y.

The modal restrictor view is Y-shaped: it can handle both indicatives and counterfactuals (Kratzer, 1981). To explain how this works, we need three ingredients: a modal base, the modal force of an operator, and an ordering.

According to Kratzer, the context of an utterance supplies a *modal base*, or a function f mapping each world w to a set of propositions that is 'held fixed' when we speculate about what might or must have been true at w. When we consider what is physically possible, the modal base might

assign to each world the laws of physics that obtain at that world, but leave out physically contingent truths. When a detective speculates about who the burglar might be, the modal base might assign to each world the detective's evidence at that world. To determine what is possible (or necessary, or likely) at a world w, we need to quantify over the possible worlds where the all of the propositions in $f(w)$ are true.[12]

Different operators are associated with different kinds of *modal force*—roughly, different kinds of quantification over possible worlds. The operators 'necessarily', 'possibly', 'it is likely that', and 'it is a good possibility that' are all associated with different modal forces. Finally, the context of utterance supplies an *ordering source* g, which lets us map each world to an ordering over worlds.[13] (One possible interpretation of this ordering is the 'closeness' ordering from Section 4.2, but there are others. Conditional and unconditional statements about what ought to happen use an ordering source that ranks worlds from most to least ideal.)

We can then say that the conditional 'Necessarily if A, then B' is true at a world w just in case B is true at all the closest A-worlds to w (according to the ordering $g(w)$) where all the propositions in $f(w)$ are true. Likewise, 'Possibly if A, then B' is true at a world w just in case B is true at some of the closest A-worlds to w (according to the ordering $g(w)$) where all the propositions in $f(w)$ are true, and similarly for other operators with other modal force. For indicative conditionals, the modal base is some piece of salient known information. For counterfactual conditionals, the modal base is empty (and thus, all possible worlds are consistent with it) while the ordering source is very rich. Kratzer's account even has the material conditional account as a special case, where the modal base maps each world w to a set of propositions true only at w, and the strict conditional as another special case, where the modal base is empty and the ordering source is completely noncommittal, invariably ranking all worlds on a par with each other.

'Bare' conditionals cause trouble for the modal restrictor view. Conditionals supposedly restrict modal operators, but where is the modal

12 Kratzer's theory could be reformulated in terms of a familiar two-place accessibility relation among worlds. We might say that world x is accessible from world w (Rwx in the usual formalism) if and only if all of the propositions in $f(w)$ are true at x. A few complications arise when the modal base maps some worlds onto inconsistent sets of propositions. Kratzer wants to say that in such cases, there are non-trivial facts about what is possible; she gives the example of a modal base that assigns to each world the set of propositions that are required by a group of Maori elders in that world (Kratzer, 1981, pp. 16-20). In one world w, the elders disagree amongst themselves, and so their requirements are inconsistent. Nonetheless, there are non-trivial facts about what is necessary at w according to the elders' requirements; Kratzer claims that the structure of the set $f(w)$ of propositions plays an essential role in determining what is necessary.

13 Kratzer's ordering source officially maps worlds to sets of propositions, which are then used to create an ordering. I omit this extra step.

operator in a conditional like 'If the lights in his study are on, then Roger is home'? Kratzer (1979, 1981) argues that conditionals without overt modal operators nonetheless contain implicit modal operators; the underlying logical form of the example conditional is '(MUST: the lights in his study are on) Roger is home'; the epistemic 'MUST' is unspoken.

Heim (1982) provides evidence for Kratzer's modalized interpretation of bare conditionals in the form of 'donkey sentences' like 'If John owns a donkey, then he feeds it carrots'. On at least one plausible reading, our sample donkey sentence means that John feeds carrots to every donkey he owns—or in more cumbersome terms, for every x such that x is a donkey and John owns x, John feeds x carrots. If the conditional were an ordinary two-place connective, we would have trouble explaining how the same variable x, bound by the same quantifier, could occur in both the antecedent and the consequent of the donkey sentence. The conditional would have the form $A \to B$, where A contained a quantifier ranging over donkeys. But Kratzer's restrictor analysis, together with the assumption that bare conditionals contain a tacit necessity operator, gives the correct reading, while providing a uniform treatment of bare and modalized conditionals.

It is often claimed that Kratzer's modal restrictor theory allows us to escape the triviality results of Section 7. Rothschild (2013), for instance, suggests that Kratzer can escape the triviality results by denying THE THESIS. To illustrate Rothschild's argument, let's consider the conditional I originally used to motivate THE THESIS.

ACE. If I draw a red ace, then it will be the ace of hearts.

I accept the conditional:

CHANCY ACE. With probability 1/2, if I draw a red ace, then it will be the ace of hearts.

Rothschild suggests that on Kratzer's account, CHANCY ACE does not express the thought that ACE has probability 1/2, or the thought that the probability of ACE's being true is 1/2. When I assert CHANCY ACE, I am not asserting that ACE has probability 1/2. Furthermore, when I am 50% confident that if I draw a red ace, it will be the ace of hearts, this does not amount to my being 50% confident that ACE is true.

Charlow (2015) argues that even if Rothschild is right, Kratzer's account is still vulnerable to the triviality result, since Equation 5 and Equation 8 can be motivated independently of THE THESIS. He goes on to argue that other easy ways out of the triviality result fail on the modal restrictor view.

10 CONCLUSION

Conditionals are important in both everyday reasoning and philosophical argument. There are conditional beliefs, conditional assertions, and conditional propositions, all of which can figure in arguments. The theories canvassed in this article try to systematize the broad range of data about which conditionals seem true, and which inferences seem valid. More phenomena remain to be explained: this article has focused on conditional beliefs and assertions, and on conditionals in English.

We can gather the similarities among the accounts discussed above into a sort of rake-shaped theory (a generalization of Bennett's concept of a Y-shaped theory), with a short 'handle' that captures what is common to all conditionals, which then splits into many 'tines' that capture the particularities of individual theories. All of the theories we have considered so far have the following commitments in common.

1. Conditionals are evaluated at 'points'.
2. To evaluate a conditional 'if A, then C' at a point p, one generates a new point q by adding the information in A to p.
3. The evaluation of the consequent C at q is the evaluation of the entire conditional at p.

The accounts disagree about the natures of points, what status conditionals and their consequents should be evaluated for, and what adding an antecedent amounts to. Table 1 summarizes how different views answer this question. (NB: Selection function and relevant logic accounts typically treat the initial point and the new point as belonging to different types—the initial point is a world, while the new point is a set of worlds. But we can ensure that both points are of the same type by rewriting the theory so that the initial point is a singleton set of one world; this is what I have done in Table 1.)

Within each of the accounts, there are open questions: the nature of the selection function; the correct interpretation of counterpossibles; how best to respond to the triviality theorems; what makes a conditional believable or assertable in a given context; how to handle bare modals on the restrictor account.

There are also open questions about how the accounts interact. Some accounts seem to be special cases of others: the past predominance view is a way of filling in the meaning of 'closest' on the closest-worlds account. At other times, different accounts appear to be rivals: it can't be both that indicative conditionals have the truth conditions given by the material interpretation, and that they lack truth values. At other times, they seem to

	Points	Status	Adding A to a Point
Strict Conditional	Sets of worlds	Truth in all worlds (original point is a singleton $\{w\}$)	Taking all worlds possible at w compatible with A
Closest Worlds and Past Predominance	Sets of worlds	Truth in all worlds (original point is a singleton $\{w\}$)	Taking all closest worlds possible at w compatible with A
Causal Modeling	Causal models with valuations	Truth in a model	Intervening to make A true
Relevant Logic	Sets of worlds	Truth in all worlds (original point is a singleton $\{w\}$)	Taking all worlds y such that $Rwxy$ for some world x compatible with A
Material Conditional	Worlds	Truth in the world	Doing nothing if A is true; moving to the 'absurd world' (where everything is true) otherwise
Probability Accounts	Probability functions	Probability $x \in [0, 1]$	Conditionalizing on A
Dynamic Test Theory	Contexts	Acceptability	Updating to accommodate an assertion of A
Modal Restrictors	Information states: modal base + ordering source	Obtaining with a given modal force	Taking all closest worlds to A in the modal base, according to the ordering source

Table 1: Theories of conditionals and their components

be modeling different domains: as with Pearl's causal modeling theory of counterfactuals and Starr's dynamic semantics theory of indicatives. Much of the interest for future research lies in understanding the interactions between the different models of conditionals.

If conditionals are useful in a wide variety of domains, from childhood development to everyday reasoning to philosophy, then conditionals are well worth studying. I have given reasons for thinking that conditionals are useful in a wide variety of domains. You may draw your own conclusions.

REFERENCES

Adams, E. W. (1970). Subjunctive and indicative conditionals. *Foundations of Language*, *6*(1), 89–94.

Adams, E. W. (1975). *The logic of conditionals: An application of probability to deductive logic*. D. Reidel.

Adams, E. W. (1988). Modus tollens revisited. *Analysis*, *48*(3), 122–128.

Aksenov, P. (2013). Stanislav Petrov: The man who may have saved the world. Retrieved, from http://www.bbc.com/news/world-europe-24280831

Alonso-Ovalle, L. (2009). Counterfactuals, correlatives, and disjunction. *Linguistics and Philosophy*, *32*(2), 207–244.

Amsel, E. & Smalley, D. (2000). Beyond really and truly: Children's counterfactual thinking about pretend possible worlds. In P. Mitchell & K. Riggs (Eds.), *Children's reasoning and the mind* (pp. 121–147). Psychology Press Ltd.

Appiah, A. (1985). *Assertion and conditionals*. Cambridge: Cambridge University Press.

Arregui, A. (2009). On similarity in counterfactuals. *Linguistics and Philosophy*, *32*(3), 245–278.

Ayer, A. (1954). Freedom and necessity. In *Philosophical essays*. London: Macmillan.

Barwise, J. & Perry, J. (1981). Situations and attitudes. *Journal of Philosophy*, *78*(11), 668–691.

Beall, J., Brady, R., Dunn, J. M., Hazen, A. P., Mares, E., Meyer, R. K., … Sylvan, R. (2012). On the ternary relation and conditionality. *Journal of Philosophical Logic*, *41*(3), 595–612.

Bennett, J. (1988). Farewell to the phlogiston theory of conditionals. *Mind*, *97*(388), 509–27.

Bennett, J. (2001). Conditionals and explanations. In A. Byrne, R. Stalnaker, & R. Wedgwood (Eds.), *Fact and value: Essays on ethics and metaphysics for judith jarvis thomson*. Cambridge, MA: MIT Press.

Bennett, J. (2003). *A Philosophical Guide to Conditionals*. Oxford University Press.

Bernstein, S. (2016). Omission impossible. *Philosophical Studies*, *173*(10), 2575–2589.

Berto, F. (2013). Impossible worlds. *The Stanford Encyclopedia of Philosophy*. Retrieved from http://plato.stanford.edu/entries/impossible-worlds/

Bjerring, J. C. (2013). On counterpossibles. *Philosophical Studies*, *168*(2), 1–27.

Bobzien, S. (2002). The development of modus ponens in antiquity: From Aristotle to the 2nd century AD. *Phronesis*, *47*(4), 359–94.

Bradley, R. (2000). Conditionals and the logic of decision. *Philosophy of Science*, *67*(3), S18–S32.

Brewka, G. (1991). *Nonmonotonic reasoning: Logical foundations of commonsense*. Cambridge: Cambridge University Press.

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

Brogaard, B. & Salerno, J. (2007). A counterfactual account of essence. *The Reasoner*, *1*(4).

Brogaard, B. & Salerno, J. (2008). Counterfactuals and context. *Analysis*, *68*(297), 39–46.

Brogaard, B. & Salerno, J. (2013). Remarks on counterpossibles. *Synthese*, *190*(4), 639–660.

Cantwell, J. (2013). Conditionals in Causal Decision Theory. *Synthese*, *190*(4), 661–679.

Charlow, N. (2015). Triviality for restrictor conditionals. *Noûs*, *49*(3), 1–32.

Choi, S. (2006). The simple vs. reformed conditional analysis of dispositions. *Synthese*, *148*(2), 369–379.

Choi, S. (2009). The conditional analysis of dispositions and the intrinsic dispositions thesis. *Philosophy and Phenomenological Research*, *78*(3), 568–590.

Collins, J., Hall, N., & Paul, L. A. (2004). Counterfactuals and causation: History, problems, and prospects. In J. Collins, N. Hall, & L. Paul (Eds.), *Causation and counterfactuals* (pp. 1–57). Cambridge, MA: The MIT Press.

Cross, C. B. (1990). Temporal necessity and the conditional. *Studia Logica*, *49*(3), 345–363.

Darwall, S. L. (1983). *Impartial reason*. Ithaca: Cornell University Press.

Dias, M. & Harris, P. [P.L.]. (1990). The influence of the imagination on reasoning by young children. *Developmental Psychology*, *8*(4), 305–318.

Díez, J. (2015). Counterfactuals, the discrimination problem and the limit assumption. *International Journal of Philosophical Studies*, *23*(1), 85–110.

Douven, I. & Verbrugge, S. (2013). The probabilities of conditionals revisited. *Cognitive Science*, *37*(4), 711–730.

Dowell, J. J. L. (2011). A flexible contextualist account of epistemic modals. *Philosophers' Imprint*, *11*(14), 1–25.

Dudman, V. (1983). Tense and time in english verb clusters of the primary pattern. *Australian Journal of Linguistics*, *3*(1), 25–44.

Dudman, V. (1984). Parsing 'if'-sentences. *Analysis*, *44*(4), 145–53.

Edgington, D. (1986). Do conditionals have truth-conditions? *Crítica*, *18*(52), 3–30.

Edgington, D. (1995). On conditionals. *Mind*, *104*(414), 235–329.

Edgington, D. (2004). Counterfactuals and the benefit of hindsight. In P. Dowe & P. Noordhof (Eds.), *Cause and chance: Causation in an indeterministic world* (pp. 12–27). Routledge.

Edgington, D. (2008). Counterfactuals. *Proceedings of the Aristotelian Society*, *108*(1), 1–21.

Egan, A. (2009). Billboards, bombs and shotgun weddings. *Synthese*, *166*(2), 251–279.

Fine, K. (1975). Critical notice of Lewis, counterfactuals. *Mind*, *84*(335), 451–458.

Galles, D. & Pearl, J. (1998). An axiomatic characterization of causal counterfactuals. *Foundations of Science*, *3*(1), 151–182.

Gärdenfors, P. (1988). *Knowledge in flux: Modeling the dynamics of epistemic states*. Cambridge, MA: MIT Press.

Gibbard, A. (1981). Two Recent Theories of Conditionals. In W. Harper, R. C. Stalnaker, & G. Pearce (Eds.), *Ifs* (pp. 211–247). Reidel.

Gibbard, A. & Harper, W. L. (1981). Counterfactuals and two kinds of expected utility. In W. Harper, R. C. Stalnaker, & G. Pearce (Eds.), *Ifs* (pp. 153–190). Reidel.

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

Gillon, B. (2011). Logic in classical Indian philosophy. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2011). Retrieved from http://plato.stanford.edu/archives/sum2011/entries/logic-india/

Gopnik, A. (2009). *The philosophical baby: What children's minds tell us about truth, love, and the meaning of life*. Random House.

Grice, H. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.

Hájek, A. (manuscript). *Most counterfactuals are false*.

Hall, N. & Hájek, A. (1994). The hypothesis of the conditional construal of conditional probability. In *Probabilities and conditionals: Belief revision and rational decision* (pp. 75–110). Cambridge: Cambridge University Press.

Harris, P. [Paul]. (2000). *The work of the imagination: Understanding children's worlds*. Blackwell Publishing.

Heim, I. (1982). *The semantics of definite and indefinite noun phrases* (Doctoral dissertation, University of Massachusetts).
Hintikka, J. (1975). Impossible possible worlds vindicated. *Journal of Philosophical Logic*, *4*(4), 475–484.
Huber, F. (2013). Structural equations and beyond. *Review of Symbolic Logic*, *6*(4), 709–732.
Jackson, F. (1987). *Conditionals*. Cambridge, MA: Blackwell Publishing.
Jago, M. (2013). Recent work in relevant logic. *Analysis*, *73*(3), 526–541.
Kaufmann, S. (2004). Conditioning against the grain. *Journal of Philosophical Logic*, *33*(6), 583–606.
Kolodny, N. & MacFarlane, J. (2010). Ifs and oughts. *Journal of Philosophy*, *107*(3), 115–143.
Koons, R. (2014). Defeasible reasoning. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2014). Retrieved from http://plato.stanford.edu/entries/reasoning-defeasible/
Krakauer, B. (2012). *Counterpossibles* (Doctoral dissertation, University of Massachusetts).
Kratzer, A. (1979). Conditional necessity and possibility. In R. Bäurle, U. Egli, & A. Stechow (Eds.), *Semantics from different points of view* (pp. 117–47). Springer.
Kratzer, A. (1981). The notional category of modality. In H. Eikmeyer & H. Reiser (Eds.), *Words, worlds, and contexts*. de Gruyter.
Kratzer, A. (2012). *Modals and conditionals: New and revised perspectives*. Oxford: Oxford University Press.
Krzyzanowska, K. (2013). Belief ascription and the Ramsey test. *Synthese*, *190*(1), 21–36.
Lauer, S. & Condoravdi, C. (2014). Preference-conditioned necessities: Detachment and practical reasoning. *Pacific Philosophical Quarterly*, *95*(4), 584–621.
Lewis, C. (1918). *Survey of symbolic logic*. University of California Press.
Lewis, D. (1973a). *Counterfactuals*. Blackwell Publishing.
Lewis, D. (1973b). Counterfactuals and comparative possibility. *Journal of Philosophical Logic*, *2*(4), 418–446.
Lewis, D. (1975). Adverbs of quantification. In E. Keenan (Ed.), *Semantics of natural language* (pp. 3–15). Cambridge: Cambridge University Press.
Lewis, D. (1976). Probabilities of conditionals and conditional probabilities. *The Philosophical Review*, *85*(3), 297–315.
Lewis, D. (1979). Counterfactual dependence and time's arrow. *Noûs*, *13*(4), 455–476.
Lewis, K. S. (2015). Elusive counterfactuals. *Noûs*, *49*(4).
Lillard, A. (2001). Pretend play as twin earth: A social-cognitive analysis. *Developmental Review*, *21*(4), 495–531.

Loewer, B. (1976). Counterfactuals with disjunctive antecedents. *Journal of Philosophy*, *73*(16), 531–537.

Mares, E. D. (1994). Why we need a relevant theory of conditionals. *Topoi*, *13*(1), 31–36.

Mares, E. D. & Fuhrmann, A. (1995). A relevant theory of conditionals. *Journal of Philosophical Logic*, *24*(6), 645–665.

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

McGee, V. (1989). Conditional probabilities and compounds of conditionals. *Philosophical Review*, *98*(4), 485–541.

McGee, V. (2005). 24.241 logic I, fall 2005. Retrieved from http://ocw.mit.edu/courses/linguistics-and-philosophy/24-241-logic-i-fall-2005/readings/chp14.pdf

Mckay, T. & Inwagen, P. V. (1977). Counterfactuals with disjunctive antecedents. *Philosophical Studies*, *31*(5), 353–356.

Melia, J. (2001). Reducing possibilities to language. *Analysis*, *61*(1), 19–29.

Menzies, P. (2014). Counterfactual theories of causation. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2014). Retrieved from http://plato.stanford.edu/archives/spr2014/entries/causation-counterfactual/

Moore, G. (1912). *Ethics*. London: Williams and Norgate.

Morgan, C. G. & Mares, E. D. (1995). Conditionals, probability, and non-triviality. *Journal of Philosophical Logic*, *24*(5), 455–467.

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

Nolan, D. (1997). Impossible worlds: A modest approach. *Notre Dame Journal of Formal Logic*, *38*(4), 535–572.

Nolan, D. (2013). Why historians (and everyone else) should care about counterfactuals. *Philosophical Studies*, *163*(2), 317–335.

Nozick, R. (1981). *Philosophical explanations*. Cambridge, MA: Harvard University Press.

Nute, D. (1975). Counterfactuals and the similarity of words. *Journal of Philosophy*, *72*(21), 773–778.

Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd). Cambridge: Cambridge University Press.

Phillips, I. (2007). Morgenbesser cases and closet determinism. *Analysis*, *67*(293), 42–49.

Pollock, J. L. (1976). The 'possible worlds' analysis of counterfactuals. *Philosophical Studies*, *29*(6), 469–476.

Prior, E. W., Pargetter, R., & Jackson, F. (1982). Three theses about dispositions. *American Philosophical Quarterly*, *19*(3), 251–257.

Ramsey, F. (1978). Law and causality. In D. Mellor (Ed.), *Foundations* (pp. 128–51). Routledge.

Reiss, J. (2009). Counterfactuals, thought experiments, and singular causal analysis in history. *Philosophy of Science*, *76*(5), 712–723.

Rothschild, D. (2013). Do indicative conditionals express propositions? *Noûs*, *47*(1), 49–68.

Routley, R. & Meyer, R. (1972a). The semantics of entailment II. *Journal of Philosophical Logic*, *1*(1), 53–73.

Routley, R. & Meyer, R. (1972b). The semantics of entailment III. *Journal of Philosophical Logic*, *1*(2), 192–208.

Routley, R. & Meyer, R. (1973). The semantics of entailment I. In H. Leblanc (Ed.), *Truth, syntax, and semantics* (pp. 194–243). North-Holland.

Ryle, G. (1950). 'If', 'so', and 'because'. In M. Black (Ed.), *Philosophical analysis*. Ithaca, NY: Cornell University Press.

Salerno, J. & Brogaard, B. (2007). Williamson on counterpossibles. *The Reasoner*, *1*(3).

Schulz, K. (2011). 'if you'd wiggled A, then B would've changed'. *Synthese*, *179*(2), 239–251.

Sider, T. (2002). The ersatz pluriverse. *The Journal of Philosophy*, *99*(6), 279–315.

Slote, M. (1978). Time in counterfactuals. *The Philosophical Review*, *87*(1), 3–27.

Sobel, J. H. (1970). Utilitarianisms: Simple and general. *Inquiry*, *13*(1-4), 394–449.

Sosa, E. (1999). How to defeat opposition to Moore. *Philosophical Perspectives*, *13*, 141–153.

Spohn, W. (2015). Conditionals: A unified ranking-theoretic perspective. *Philosophers' Imprint*, *15*(1).

Stalnaker, R. (1968). A theory of conditionals. *American Philosophical Quarterly*, 98–112.

Stalnaker, R. (1975). Indicative conditionals. *Philosophia*, *5*(3), 269–86.

Stalnaker, R. (1981). A defense of conditional excluded middle. In W. Harper, R. C. Stalnaker, & G. Pearce (Eds.), *Ifs* (pp. 87–104). D. Reidel.

Stalnaker, R. (1999). Assertion. In *Context and content*. Oxford University Press.

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

Thomason, R. & Gupta, A. (1980). A theory of conditionals in the context of branching time. *Philosophical Review*, *89*(1), 65–90.

Vinci, T. C. (1988). Objective chance, indicative conditionals and decision theory; or, how you can be smart, rich and keep on smoking. *Synthese*, *75*(1), 83–105.

von Fintel, K. (2001). Counterfactuals in a dynamic context. In M. Kentstowicz (Ed.), *Ken Hale: A life in language*. Cambridge, MA: MIT Press.

von Fintel, K. (2011). Conditionals. In K. von Heusinger, C. Maienborn, & P. Portner (Eds.), *Semantics: An international handbook of meaning* (pp. 1515–1538). DeGruyter.

Walton, D. (2001). Are some modus ponens arguments deductively invalid? *Informal Logic*, *22*(1), 19–46.

Walton, K. (1990). *Mimesis as make-believe: On the foundations of the representational arts*. Harvard University Press.

Warmbrod, K. (1982). A defense of the limit assumption. *Philosophical Studies*, *42*(1), 53–66.

Weisberg, D. & Gopnik, A. (2013). Pretense, counterfactuals, and bayesian causal models: Why what is not real really matters. *Cognitive Science*, *37*(7), 1368–1381.

Williams, J. R. G. (2008). Conversation and conditionals. *Philosophical Studies*, *138*(2), 211–223.

Williams, J. R. G. (2012). Counterfactual triviality: A Lewis-impossibility argument for counterfactuals. *Philosophy and Phenomenological Research*, *3*(85), 648–670.

Williamson, T. (2007). *The philosophy of philosophy*. Oxford: Blackwell Publishing.

Woodward, J. (2004). Counterfactuals and causal explanation. *International Studies in the Philosophy of Science*, *18*(1), 41–72.

Yagisawa, T. (2010). *Worlds and individuals, possible and otherwise*. Oxford University Press.

Yalcin, S. (2007). Epistemic modals. *Mind*, *116*(464), 983–1026.

Yalcin, S. (2012a). A counterexample to modus tollens. *Journal of Philosophical Logic*, *41*(6), 1001–1024.

Yalcin, S. (2012b). Bayesian expressivism. *Proceedings of the Aristotelian Society*, *112*(2), 123–160.