

Indicative and Subjunctive Conditionals

Brian Weatherson

2001

Abstract: In any plausible semantics for conditionals, the semantics for indicatives and subjunctives will resemble each other closely. This means that if we are to keep the possible-worlds semantics for subjunctives suggested by Lewis, we need to find a possible-worlds semantics for indicatives. One reason for thinking that this will be impossible is the behaviour of rigid designators in indicatives. An indicative like ‘If the stuff in the rivers, lakes and oceans really is H_2O , then water is H_2O ’ is non-vacuously true, even though its consequent is true in no possible worlds, and hence not in the nearest possible world where the antecedent is true. I solve this difficulty by providing a semantics for conditionals within the framework of two-dimensional modal logic. In doing so, I show that we can have a reasonably unified semantics for indicative and subjunctive conditionals.

THIS paper presents a new theory of the truth conditions for indicative conditionals. The theory allows us to give a fairly unified account of the semantics for indicative and subjunctive conditionals, though there remains a distinction between the two classes. Put simply, the idea behind the theory is that the distinction between the indicative and the subjunctive parallels the distinction between the necessary and the *a priori*. Since that distinction is best understood formally using the resources of two-dimensional modal logic, those resources will be brought to bear on the logic of conditionals.

1 A Grand Unified Theory?

Our primary focus is the indicative conditional ‘If A , B ’, written as $A \rightarrow B$. Most theorists fail to distinguish between this conditional and ‘If A , then B ’, and for the most part I will follow this tradition. The most notable philosophical exception is Grice, who suggested that only the latter says that B follows from A in some relevant way (1989, 63). Theorists do distinguish between this conditional and the subjunctive ‘If it were the case that A , it would be the case that B ’, written as $A \Box \rightarrow B$. There is some debate about precisely where to draw the line between these two classes, which I’ll discuss in section three, but for now I’ll focus on cases far from the borderline. One important tradition in work on conditionals holds that the semantics of indicatives differs radically from the semantics of subjunctives. According to David Lewis (1973, 1976) and

Frank Jackson (1987) for example, indicatives are truth-functional, but subjunctives are not. This makes a mystery of some of the data. For example, as Jackson himself writes:

Before the last presidential election commentators said ‘If Reagan loses, the opinion polls will be totally discredited’, afterwards they said ‘If Reagan had lost, the opinion polls would have been totally discredited’, and this switch from indicative to subjunctive counterfactual did *not* count as a change of mind (Jackson 1987, 66).

The point can be pushed further. To communicate the commentators’ pre-election opinions using indirect speech we would say something like (1).

- (1)** Commentators have said that if Reagan were to lose the opinion polls would be totally discredited.

Yet it is possible on Jackson’s view that what the commentators said was *true*, since Reagan won, yet the words after ‘that’ in (1) form a *false* sentence. So we can accurately report someone speaking truly by using a false sentence. Jackson’s response plays on the connections between $A \rightarrow B$ and the disjunction ‘Not- A or B ’. That disjunction has undeniably different truth conditions to $A \Box \rightarrow B$. Pushing the truth conditions of $A \rightarrow B$ closer to those of $A \Box \rightarrow B$ will move them away from ‘Not- A or B ’. One gain in similarity and theoretical simplicity is bought at the cost of another. Jackson’s account, by making $A \rightarrow B$ have similar truth conditions to ‘Not- A or B ’ but similar assertibility conditions to $A \Box \rightarrow B$, tries to have the best of both worlds. How great the similarity between indicative conditionals and disjunctions really is, and hence how great the cost of linking indicatives and subjunctives, might well be questioned. After all, we don’t report an utterance of an indicative using a disjunction.

Two types of cases seem to threaten the success of a unified theory. First, rigidifying expressions like ‘actually’ behave differently in indicatives and subjunctives. Secondly, some conditionals differ in intuitive truth value when we transpose them from the indicative to the subjunctive. The most famous examples of this phenomenon involve various presidential assassinations. The effects of rigidity on conditionals are less explored, so we will first look at that. Consider the following example, from page 55 of *Naming and Necessity*.

- (2)** If heat had been applied to this stick S at t_0 , then at t_0 stick S would not have been one meter long.

The background is that we have stipulated that a metre is the length of stick S at time t_0 . (2) contrasts with (3), which seems false.

(3) If heat was applied to this stick S at t_0 , then at t_0 stick S was not one meter long.

If we have stipulated that to be a meter long is to be the length of S at t_0 , then whatever conditions S was under at t_0 , it was one meter long. As Jackson points out, we can get the same effect with explicit rigidifiers like ‘actually’. We could, somewhat wistfully, say

(4) If Hillary Clinton were to become the next U.S. President, things would be different from the way they actually will be.

(5) If Hillary Clinton becomes the next U.S. President, things will be different from the way they actually will be.

It looks like any theory of conditionals will have to account for a difference between the behaviour of rigid designators in indicatives and subjunctives. We may avoid the conclusion by showing that the difference only appears in certain types of conditionals, and we already have an explanation for those cases. For example, it is well known that usually one cannot say $A \rightarrow B$ if it is known that not- A . As Dudman (1994) points out, **(6)** is clearly infelicitous on its most obvious reading.

(6) *Granny won, but if she lost she was furious.

To complete the diagnosis, note that the most striking examples of the different behaviour of rigid designators in different types of conditionals comes up in cases where the antecedent is almost certainly false. The effect is that the subjunctive can be asserted, but not the indicative. So this phenomenon may be explainable by some other part of the theory of conditionals.¹ These are the most striking exemplars of the difference I am highlighting, but not the only examples. Hence, this point cannot explain all the data, though it may explain why pairs like **(2)/(3)** and **(4)/(5)** are striking. For instance, in the following pairs, the indicative seems appropriate and intuitively true, and the subjunctive seems inappropriate and intuitively false.

(7) If C-fibres firing is what causes pain sensations, then C-fibres firing is what actually causes pain sensations.

(8) If C-fibres firing were what caused pain sensations, then C-fibres firing would be what actually causes pain sensations.

(9) If the stuff that plays the gold role has atomic number 42, then gold has atomic number 42.

¹An anonymous reviewer for *Philosophical Quarterly* suggested this point.

- (10)** If the stuff that played the gold role had atomic number 42, gold would have atomic number 42.

In (9) and (10) I assume that to play the gold role one must play it throughout a large part of the world, and not just on a small stage. Something may play the gold role in a small part of the world without being gold. Since there are pairs of conditionals like these where the indicative is appropriate, but the subjunctive is not, the explanation of the behaviour of rigid terms cannot rely on the fact that the antecedents of indicatives must be not known to be false. We will also need a more traditional example of the differences between indicatives and subjunctives, as in (11) and (12).

- (11)** If Hinckley didn't shoot Reagan, someone else did.
(12) If Hinckley hadn't shot Reagan, someone else would have.

I have concentrated on the examples involving rigidity because they seem to pose a deeper problem for unifying the theory of conditionals than the presidential examples. As Jackson (1987, 75) points out, one can presumably explain (11) and (12) on a possible worlds account by varying the similarity metric between indicatives and subjunctives, or on a probabilistic account by varying the background evidence. It is unclear, however, how this will help with the rigidity examples. Assume, for example, that C-fibres firing is not what causes pain sensations. Still, (7) seems true, but its consequent is false in all possible worlds. Therefore, the nearest world in which its antecedent is true is a world in which its consequent is false, and on a simple possible worlds theory it should turn out false. On a simple probabilistic account, the probability that C-fibres firing actually cause pain sensations given that they do is 1, whatever the background evidence, so (8) should turn out true, contrary to our intuitions. So while the details deal with the presidential examples, the structure of the theory must deal with the rigidity examples.

I will follow that strategy here. In section two I set out the framework of a unified possible worlds account of indicatives and subjunctives. In section three I present my preferred way of filling out the details of that framework. The framework deals with the differing behaviour of rigid designators in indicatives and subjunctives; the details deal with examples like (11) and (12). One reason for dividing the presentation in this way is to highlight the option of accepting the framework and filling in the details in different ways.

2 The New Theory

2.1 Actually

As Kripke ([1972] 1980) showed, the reference for some terms is fixed by what plays a particular role in the actual world. Even if it were the case that XYZ fills the ocean, falls from the sky, is drinkable and transparent and so on, for short is *watery*, it would still be the case that water is H₂O, not XYZ. For it would still be that H₂O *actually* is watery. Whatever were the case, this world would be actual.

Yet, we want to have a way to talk about what would have happened had some other world been actual. In particular, had the actual world been one in which XYZ is watery, it would be true, indeed necessarily true, that water is XYZ. Throughout the 1970s a number of methods for doing this were produced. The following presentation is indebted to Davies and Humberstone (1980), but other approaches might have been used. The notation $\models_y^x A$ is interpreted as ‘ A is true in world y from the perspective of world x as actual’. So, letting @ be the actual world and W be a world in which only XYZ is watery, we can represent what was said informally above as follows.

$\models_{@} H_2O$ is watery and H₂O is water.

$\models_w^@ XYZ$ is watery and H₂O is water.

$\models_{@}^w H_2O$ is watery and XYZ is water.

$\models_w^w XYZ$ is watery and XYZ is water.

Now as Kripke noted, it is necessary but *a posteriori* that water is H₂O. Conversely, it is *a priori* but contingent that water is watery. This is *a priori* because we knew before we determined what water really is that it would be whatever plays the watery role in this world, the actual world. In general A is necessary iff, given this is the actual world, it is true in all worlds. And A is *a priori* iff, whatever the actual world turns out to be like, it makes A true. So we get the following definitions.

A is *a priori* iff for all worlds W , $\models_w^w A$.

A is necessary iff for all worlds W , $\models_w^@ A$.

The connection between actuality and the *a priori* is important. It is *a priori* that we are in the actual world. Something is *a priori* iff it is true whenever the two indices are the same. If we regard possible worlds as sets of sentences, we can think of the sets $\{A: \models_x^x A\}$ for each possible world x as the epistemically possible worlds. Note that I don’t make the set of epistemically possible worlds relative to an evidence set, as others commonly do. Rather they are just the sets of sentences consistent with what we know

a priori. More accurately, identify a world pair $\langle x, y \rangle$ with the set of $\{A: \models_y^x A\}$. Then $\langle x, y \rangle$ is an epistemically possible world pair iff $x = y$.

To finish this formal excursion, we note the definition of ‘Actually A ’. Given what has been said so far, this needs no explanation.

$$\models_y^x \text{Actually } A \text{ iff } \models_x^x A.$$

2.2 The Analysis of Indicatives

Now we have the resources for my theory of the truth conditions for indicatives. I also give the parallel truth condition for subjunctives to show the similarities.

$$\models_{@}^@ A \rightarrow B \text{ iff the nearest possible world } x \text{ that } \models_x^x A \text{ is such that } \models_x^x B.$$

$$\models_{@}^@ A \Box \rightarrow B \text{ iff the nearest possible world } x \text{ that } \models_x^@ A \text{ is such that } \models_x^@ B.$$

These only cover the special case of what is true here from the perspective of this world as actual. We can partially generalise the analysis of indicatives in one dimension as follows.

$$\begin{aligned} \models_w^w A \rightarrow B \text{ iff the nearest possible world } x \text{ to } W \text{ such that } \models_x^x A \text{ is such that} \\ \models_x^x B. \end{aligned}$$

I will make some comments below about how we might fully generalise the analysis, but for now, I want to focus on these simpler cases. Note that straight away this makes $A \rightarrow \text{Actually } A$ come out true, by the definition of ‘Actually’. If we allow ourselves quantification over propositions, we can give an analysis of ‘things are different from the way they actually are’, as follows:

$$\begin{aligned} (\models_y^x \text{ Things are different from the way they actually are}) \text{ iff} \\ (\exists p: \models_y^x p \text{ and not } \models_x^x p) \end{aligned}$$

Since nothing both is and is not the case in x from the perspective of x as actual, this can never be true when y is x . This explains why it can never serve as the consequent of an indicative conditional.

2.3 Motivations

The theory outlined here is reasonably unified, and accounts for the rigidity phenomena, but without any further justification, the resort to two-dimensional modal logic is *ad hoc*. This subsection responds to that problem with some independent motivations for the theory. In particular I argue that this theory best captures the well-known epistemic feel of the indicative conditional.

Ever since Ramsey (1929/1990) most theorists have held that there is an epistemic element to indicatives. Here is Ramsey's sketch of an analysis of indicatives.

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 , $\neg q$ ' are contradictories (Ramsey 1929/1990, 247n).

Nothing of the sort could be true about subjunctives. What is in our 'stock of knowledge', or the contextually relevant knowledge, makes at most an indirect contribution to the truth-value of a subjunctive. It makes an indirect contribution because the common knowledge might affect the context, which in turn determines the similarity measure. But given a context, a subjunctive makes a broadly metaphysical claim, an indicative a broadly epistemic claim. Hence, the relationship between the indicative and subjunctive should parallel the relationship between the necessary and the *a priori*. As should be clear, this is exactly what happens on this theory.

The close similarity between the indicative/subjunctive distinction and the *a priori*/necessary distinction can be demonstrated in other ways. For example, corresponding to the contingent *a priori* (13) the indicative (14) is true, but the subjunctive (15) is false. And corresponding to the necessary *a posteriori* (16) the subjunctive (17) is true but the indicative (18) is false. (I am assuming that it is part of the definitions of the water role and the fire role that nothing can play both roles.)

- (13) Water is what plays the water role.
- (14) If XYZ plays the water role, XYZ is water.
- (15) If XYZ played the water role, it would be water.
- (16) Water is H₂O.
- (17) If all H₂O played the fire role, all water would be fire.
- (18) If all H₂O plays the fire role, all water is fire.

This suggests the analysis sketched here is not *ad hoc* at all, but follows naturally from considerations about the necessary and *a priori*. These sketchy considerations might not provide much positive support for my theory. The main evidence for the theory, however, is the way it manages the hard cases, particularly cases involving rigid designation. What these considerations show is that the correct theory of indicatives may invoke the resources of two-dimensional modal logic without automatically renouncing any claim to systematicity.

3 The Details

In this section, I want to look at four questions. First, what can we say about the similarity measure at the core of this account? Secondly, how should we generalise the theory to cover cases where the definite description in the analysis appears to denote nothing? Thirdly, how should we generalise the theory to cover cases where the two indices differ? Finally, how should we draw the line between indicatives and subjunctives? If what I said in the previous section is correct, there should be something to say about each of these questions, and what is said should be motivated. While it is not important that what I say here is precisely true, I do hope that it is.

3.1 Nearness

Ideally, we could use exactly the same similarity metric for both indicatives and subjunctives. The existence of pairs like (11) and (12) suggests this is impossible. So we must come up with a pair of measures on the worlds satisfying three constraints. First, the measure for subjunctives must deliver plausible verdicts for most subjunctive conditionals. Secondly, the measure for indicatives must deliver plausible verdicts for most indicative conditionals. Thirdly, the measures must be similar enough that we can explain the close relationship between indicatives and subjunctives set out in section one. The theory of section two requires that these objectives be jointly satisfiable. I will attempt to demonstrate that they are by outlining a pair of measures satisfying all three.

Lewis (1979a) provides the measure for subjunctives. He suggests the following four rules for locating the nearest possible world in which A is true.

1. It is of the first importance to avoid big, widespread, diverse violations of law.
2. It is of the second importance to maximise the spatio-temporal region throughout which perfect match of particular fact prevails.
3. It is of the third importance to avoid even small, localized, simple violations of law.
4. It is of little or no importance to secure approximate similarity of particular fact, even in matters that concern us greatly. (Lewis 1979a, 47–48)

The right measure for indicatives is somewhat simpler. Notice that whenever we know that $A \supset B$ and don't know whether A , $A \rightarrow B$ seems true. More generally, if I know some sentence S such that A and S together entail B , and I would continue to know S even were I to come to doubt B , then $A \rightarrow B$ will seem true to me. No matter how good a card cheat I know Sly Pete to be, if I know that he has the worse hand, and that whenever someone with the worse hand calls they lose, it will seem true to me that *If*

Sly Pete calls, he will lose. Further, if someone else knows these background facts and tells me that *If Sly Pete calls, he will lose*, she speaks truthfully.

This data suggests that whenever there is a true S such that A and S entail B , $A \rightarrow B$ is true. But this would mean $A \rightarrow B$ is true whenever $A \supset B$ is true, which seems incredible. On this theory it is true that *If there is a nuclear war tomorrow, life will go on as normal.* There are some very subtle attempts to make this palatable. The ‘Supplemented Equivalence Theory’ in Jackson (1987) may even be successful. But two problems remain for all theories saying $A \rightarrow B$ has the truth value of $A \supset B$. First, they make some apparently true negated conditionals turn out false, such as *It is not true that if there is a nuclear war tomorrow, life will go on as normal.* It is hard to see how an appeal to Gricean pragmatics will avoid this problem. Secondly, such theories fail the third task we set ourselves at the start of the section: explaining the close connections between indicatives and subjunctives.

So we might be tempted to try a different path. Let’s take the data at face value and say that $A \rightarrow B$ is true in a context if there is some S such that some person in that context knows S , and A and S together entail B . We can formalise this claim as follows. Let $d(x, y)$ be the ‘distance’ from x to y . This function will satisfy few of the formal properties of a distance relationship, so remember this is just an analogy. Let \mathbf{K} be the set of all propositions S known by someone in the context, \mathcal{W} the set of all possible worlds, and i the impossible world, where everything is true. Then $d: \mathcal{W} \times \mathcal{W} \cup i \rightarrow \mathbb{R}$ is as follows:

If $y = x$ then $d(x, y) = 0$

If $y \in \mathcal{W}$, $y \neq x$ and $\forall S: S \in \mathbf{K} \supset \models_y S$, $\forall S: S \in \mathbf{K} \supset \models_y S$, then $d(x, y) = 1$

If $y = i$ then $d(x, y) = 2$

Otherwise, $d(x, y) = 3$

Less formally, the nearest world to a world is itself. The next closest worlds are any compatible with everything known in the context, then the impossible world, then the possible worlds incompatible with something known in the context. It may seem odd to have the impossible world closer than some possible worlds, but there are two reasons for doing this. First, in the impossible world everything known to any conversational participant is true. Secondly, putting the impossible world at this position accounts for some examples. This is a variant on a well known case; see for example Gibbard (1981) and Barker (1997).

Jack and Jill are trying to find out how their local representative Kim, a Democrat from Texas, voted on a resolution at a particular committee meeting. So far, they have not even found out whether Kim was at the meeting. Jack finds out that all Democrats

at the meeting voted against the resolution; Jill finds out that all Texans at the meeting voted for it. When they return to compare notes, Jack can truly say *If Kim was at the meeting, she voted against the resolution*, and Jill can truly say *If Kim was at the meeting, she voted for the resolution*. If i is further from the actual world than some possible world where Kim attended the meeting, these statements cannot both be true.

It may be thought the distance function needs to be more fine-grained to account for the following phenomena². It seems possible that in each of the following pairs, the first sentence is true and the second false.

(19)

- a. If Anne goes to the party, so will Billy.
- b. If Anne goes to the party, Billy will not go.

(20)

- a. If Anne and Carly go to the party, Billy will not go.
- b. If Anne and Carly go to the party, so will Billy.

(21)

- a. If Anne, Carly and Donna go to the party, so will Billy.
- b. If Anne, Carly and Donna go to the party, Billy will not.

Assume, as seems plausible, it is necessary and sufficient for $A \rightarrow B$ to be true that the nearest $A \wedge B$ world is closer than the nearest $A \wedge \neg B$ world. (This does not immediately follow from the analysis in section 2, but is obviously compatible with it.) Given this, there is no context in which the first conditional in each pair is true, and the second false. McCawley (1996) points out a way to accommodate these intuitions. Every time a conditional is uttered, or considered in a private context, the context shifts so as to accommodate the possibility that its antecedent is true. So at first we don't consider worlds where Carly or Donna turn up, and agree that (19a) is true and (19b) false because in those worlds Billy loyally follows Anne to the party. When (20a) or (20b) is uttered, or considered, we have to allow some worlds where Carly goes to the party into the context set. In some of these worlds Anne goes to the party and Billy doesn't, the worlds where Carly goes to party. A similar story explains how (21a) can be true despite (20b) being false.³

²Lewis (1973) makes this objection to a similar proposal for subjunctives; the objection has just as much force here as it does in the original case.

³There is an obvious similarity between this argument and some of the uses of contextual dependence in Lewis's theory of knowledge (Lewis 1996). Indeed, McCawley credits Lewis (1979b) as an inspiration for his ideas.

This move does seem to save the theory from potentially troubling data, but without further support it may seem rather desperate. There are two independent motivations for it. First, it explains the inappropriateness of (6).

(6) *Grannie won, but if she lost she was furious.

If assertion narrows the contextually relevant worlds to those where the assertion is true, as Stalnaker (1978) suggests, and uttering a conditional requires expanding the context to include worlds where the antecedent is true, it follows that utterances like (6) will be defective. The speech acts performed by uttering each clause give the hearer opposite instructions regarding how to amend the context set. Secondly, McCawley's assumption explains why we generally have little use for indicative conditionals whose antecedents we know are false. To interpret an indicative we first have to expand the context set to include a world where the antecedent is true, but if we know the antecedent is false we usually have little reason to want to do that. If there is a dispute over the size of the context set, we may want to expand it so as to avoid miscommunication, which explains why we will sometimes assert conditionals with antecedents we know to be false when trying to convince someone else that the antecedent really is false.

So we have a pair of measures that give plausible answers on a wide range of cases. Such a pair should also validate the close connection between indicatives and subjunctives we saw earlier. The data set out in section one suggests that this connection may be close to synonymy, as in (1), but in some cases, as in (11) and (12), the connection is much looser. The differing behaviour of rigid designators in indicatives and subjunctives reveals a further difference, but the two-dimensional nature of the analysis, not the particulars of the similarity metric, accounts for that. I propose to explain the data by looking at which facts we hold fixed when trying to determine the nearest possible world. The facts we hold fixed in evaluating indicatives and subjunctives, according to the two metrics outlined above, are the same in just the cases we feel that the indicatives and subjunctives say the same thing.

When evaluating an indicative we hold fixed all the facts known by any member of the conversation. When evaluating a subjunctive we hold fixed (a) all facts about the world up to some salient time t and (b) the holding of the laws of nature at all times after t . The time t is the latest time such that some worlds fitting this description make A true and contain no large miracles. The two sets of facts held fixed match when we know all the salient facts about times before t , and know no particular facts about what happens after t .

In the opinion poll case, when evaluating the original indicative our knowledge at the earlier time was held fixed. We knew that the polls predicted a Reagan landslide, that when one makes spectacularly false predictions one is discredited, and so on. When

we turn to evaluating the subjunctive, we hold fixed the facts about the world before the election (presumably the relevant time t) and some laws. Therefore, we hold fixed the polls predictions, and the law that when one makes spectacularly false predictions one is discredited. So the same facts are held fixed. And in general, this will happen whenever all we know is all the specific facts up to the relevant time, and some laws that allow us to extrapolate from those facts.

In the case where indicatives and subjunctives come apart, as in (11) and (12), the relevant knowledge differs from the first case. By hypothesis, we do not know who pulled the trigger, but we do know that a trigger was pulled. Our knowledge of the relevant facts does not consist in knowledge of all the details up to a salient time, and knowledge that the world will continue in a law-governed way after this. Therefore, we would predict that the indicatives and subjunctives would come apart, because what is held fixed when evaluating the two conditionals differs. We find exactly that. So the pair of measures can explain the close connection between indicatives and subjunctives when it exists, and explain why the two come apart when they do come apart.

3.2 No Nearest Possible World

Generally, there are three kinds of problems under this heading. First, there may be no A -worlds, and so no nearest A -world. Secondly, there may be an infinite sequence of ever-nearer A -worlds without a nearest A -world. Thirdly, there may be several worlds in a tie for nearest A -world. If the measure suggested in the previous section is correct, the first two problems do not arise here. The third problem, however, arises almost all the time, so we need to say something about it.

The approach I favour is set out in Stalnaker (1981). The comparative similarity measure is a partial order on the possible worlds. Stalnaker recommends we assess conditionals using supervaluations, taking the precisifications to be the complete extensions of this partial order. In particular, if several possible worlds tie for being the closest A -worlds⁴, then $A \rightarrow B$ will be true if they are all B -worlds, false if they are all $\neg B$ -worlds, and not truth-valued otherwise. For consistent A , this makes $\neg(A \rightarrow B)$ equivalent to $A \rightarrow \neg B$. Since we generally deny $A \rightarrow B$ just when we would be prepared to assert $A \rightarrow \neg B$, this seems like a good outcome.⁵ Further, this account makes $A \rightarrow B$ generally come out gappy when A is false. Many theorists hold that indicative conditionals, especially those with false antecedents, lack truth values.⁶ This can't be right in general,

⁴Of course in this context x is an A -world iff $\models_x A$.

⁵Edgington (1996) furnishes some nice examples against the view that $A \Box \rightarrow B$ should be false when there are several equally close A -worlds in a tie for closest and some are B -worlds but some are $\neg B$ -worlds.

⁶See Edgington (1995) for an endorsement of this position and discussion of others who have held it.

since it is a platitude that $A \rightarrow A$ is true for every A , but the position has some attraction. Happily, our theory respects the motivations behind such positions without violating the platitude.

In any case, these details are not important to the overall analysis. If someone favours a resolution of ties along the lines Lewis suggested this could easily be appended onto the basic theory.

3.3 The General Theory

So far, I have just defined what it is for $A \rightarrow B$ to be true in this world from the perspective of this world as actual. To have a fully general theory I need to say when $A \rightarrow B$ is true in an arbitrary world from the perspective of another (possibly different) world as actual. And that general theory must yield the theory above as a special case when applied to our world. As with the special theory above, the general theory will mostly be derived from Twin Earth considerations.

In general, $\models_y^x A \rightarrow B$ iff the nearest world pair $\langle z, v \rangle$ such that $\models_v^z A$ is such that $\models_v^z B$. Nearness is again defined epistemically, but what *we* know about x and y matters. In particular if $\models_v^z C$ for all sentences C such that someone in the context knows that $\models_y^x C$, but not $\models_w^u C$ for some such C , then $\langle z, v \rangle$ is closer to $\langle x, y \rangle$ than is $\langle u, w \rangle$. As should be clear from this, nearness is context-dependent, and the context it depends on is the actual speaker's context. For conditionals as for quantified sentences, the same words will express different propositions in different contexts.

Let's draw out some consequences of this definition. First, for any x we know that $\models_x^x C$ for all *a priori* propositions C . In particular, we know that $\models_x^x d \leftrightarrow (\text{Actually } d)$ for any proposition d , where ' \leftrightarrow ' represents the material biconditional. So the nearest world pair $\langle z, v \rangle$ to $\langle x, x \rangle$ must be one in which $z = v$, *even if that means z^* is the impossible world i* . Hence the general theory of indicatives reduces to the special theory set out above when applied to epistemically possible worlds: when assessing the truth value of an indicative in an epistemically possible world pair we need only look at other epistemically possible world pairs.

Secondly, when evaluating conditionals with respect to epistemically impossible world pairs $\langle x, y \rangle$, we need to use other epistemically impossible world pairs. For example, imagine some explorers are wandering around Twin Australia, a dry continent to the south of Twin Earth. As explorers of such lands are wont to do, they are dying of thirst, so they are seeking some watery stuff to save themselves. Without knowing whether they succeed, we know (22) is false.

(22) If the explorers find some watery stuff, they will find some water.

This theory can explain the falsity of (22). We know, from the way Twin Earth is stipulated, that all the watery stuff of the explorers' acquaintance is not water. So we know any watery stuff they find will not be water. And we know that water is scarce on Twin Earth, even scarcer than watery stuff in Twin Australia, so it is unlikely they will find some watery stuff and simultaneously stumble across some water.

This theory also explains occurrences of indicatives embedded in subjunctives. These are very odd, as should be expected if indicatives are about epistemic connections and subjunctives about metaphysical connections, but we can just make sense of them some of the time. For example, it seems possible to make sense of (23) and that it is true.

(23) If the bullet that actually killed JFK had instead killed Jackie Kennedy, then it would be true that if Oswald didn't kill Jackie Kennedy, someone else did.

On our theory, to evaluate this we first find the nearest world pair $\langle @, w \rangle$ such that $\models_w^@$ The bullet that actually killed JFK instead killed Jackie Kennedy, and then evaluate the indicative relative to it. Now one thing we know about this world pair is that in it, someone killed Jackie Kennedy. So this must hold in all nearby world pairs. Hence in any such world pair that Oswald did not kill Jackie Kennedy, someone else did, so (23) turns out true.

It might be thought that such embeddings do not make particularly good sense. I have some sympathy for such a view. If one adopts the 'special theory' developed in the previous section, and rejects the general theory developed in this subsection, one may have an explanation for the impossibility of such embeddings. However, even if we cannot make sense of such embeddings, we still need to account for the truth conditions of indicatives relative to epistemically impossible world pairs to make sense of claims such as *Necessarily* ($A \rightarrow A$).⁷

3.4 Classifying Conditionals

In recent years, there has been extensive debate over where the line between indicatives and subjunctives falls. This debate focuses on whether 'future indicatives' like (24) are properly classified with indicatives or subjunctives.

(24) If Booth doesn't shoot Lincoln, someone else will.

Jackson (1990) and Bennett (1995) argue that this should go with ordinary indicatives. Dudman (1994) and Bennett (1988) argue that it should go with ordinary subjunctives,

⁷I am indebted to Lloyd Humberstone for pointing this out to me.

though this is not how Dudman would put it. This theory of indicatives appears to favour Jackson and (the later) Bennett, because of the apparent triviality of conditionals like (25).

(25) If it will rain then it will actually rain.

4 Conclusion

Despite its lack of attention in the literature, data about the role of rigid designators in indicatives deserve close attention. Any plausible theory of indicatives must be able to deal with it, and it isn't clear how existing possible worlds theories could do so. The easiest way to build a semantics for indicatives is to say that "If A then C " is true just in case the nearest world in which A is true is a world where C is true. Even before the hard questions about the meaning of 'nearest' here start to be asked, we know a theory of this form is wrong because it makes mistaken predictions about the role of rigid designators. A conditional like "If the stuff in the rivers, lakes and oceans really is XYZ, then water is XYZ" is *true*, even though the consequent is true in no possible worlds. The simplest way to solve this difficulty is to revisit the idea of 'true in a world'. Rather than looking for a nearby world in which A is true, and asking whether C is true in it, we look for a nearby world W such that A is true under the supposition that W is actual, and ask whether C is true under the supposition that W is actual. In the terminology of Jackson (1998), we look at worlds considered as actual, rather than worlds considered as counterfactual. This simple change makes an important difference to the way rigid designators behave. There is no world in which water is XYZ. However, under the supposition that the stuff in the rivers, lakes and oceans really is XYZ, and the H₂O theory is just a giant mistake, that is, under the supposition that we are in the world known as Twin Earth, water is XYZ. In short, "water is XYZ" is true in Twin Earth *considered as actual*, even though it is false in Twin Earth considered as counterfactual. So the data about behaviour of rigid designators in indicatives, data like the truth of "If the stuff in the rivers, lakes and oceans really is XYZ, then water is XYZ", does not refute the hypothesis that "If A then C " is true iff the nearest world such that A is true in that world considered as actual is a world where C is true in that world considered as actual.

In section two we looked at how the formal structure of a theory built around that hypothesis might look. In section three we looked at how some of the details may be filled in. The most pressing task is to provide a similarity metric so we can have some idea about which worlds will count as being nearby. The theory I defended has three important features. First, it is *epistemic*. Which worlds are nearby depends on what is known by conversational participants. Secondly, it is *contextualist* in two respects. The

first respect is that it is the knowledge of the audience that matters, not just the knowledge of the speaker and the intended audience. The second respect is that it allows that what is known by the audience may be affected by the utterance of the conditional. In particular, if the utterance of “If A , B ” causes the audience to consider A to be possible, and hence cease to know that $\neg A$, then A is not part of what is known for purposes of determining which worlds are nearby. (I assume here a broadly contextualist account of knowledge, as in Lewis (1996), but this is inessential. If you do not like Lewis’s theory, replace all references to knowledge here, and in section 3.1, with references to epistemic certainty. I presume that what is epistemically certain really is contextually variable in the way Lewis suggests.) Thirdly, it is *coarse-grained*: whether a world is nearby depends only on whether it is consistent with what is known, not ‘how much’ it agrees with what is known. The resultant theory seems to capture all the data, to explain the generally close connection between indicatives and subjunctives, and to explain the few differences which do arise between indicatives and subjunctives.

The other detail to be filled in concerns embeddings of indicatives inside subjunctives. The formalism here requires that we use the full resources of two-dimensional modal logic, but the basic idea is very simple. Consider a sentence of the form “If it were the case that A , it would be the case that if B , C .” Roughly, this will be true iff the metaphysically nearest world in which A is true, call it w_A , is a world where $B \rightarrow C$ is true. And that will be true iff the epistemically nearest world to w_A in which B is true is a world where C is true. Less roughly, we have to quantify not over worlds, but over pairs of worlds, where the first element of the pair determines the reference for rigid designators, and the second element determines the truth of sentences given those references. But this only adds to the formal complexity; the underlying idea is still the same. The important philosophical point to note is that when we are trying to find the epistemically nearest world to w_A (or, more strictly, the nearest world pair to $\langle @, w_A \rangle$) the facts that have to be held fixed are the facts that we know about w_A , not what our counterparts in w_A , or indeed what any inhabitant of w_A knows about their world. These embeddings may be rare in everyday speech, but since they are our best guide to the truth values of indicatives in other possible worlds, they are theoretically very important.

References

- Barker, Stephen. 1997. “Material Implication and General Indicative Conditionals.” *The Philosophical Quarterly* 47 (187): 195–211. doi: 10.1111/1467-9213.00055.
- Bennett, Jonathan. 1988. “Farewell to the Phlogiston Theory of Conditionals.” *Mind* 97 (388): 509–27. doi: 10.1093/mind/xcvii.388.509.
- . 1995. “Classifying Conditionals: The Traditional Way Is Right.” *Mind* 104

- (414): 331–54. doi: 10.1093/mind/104.414.331.
- Davies, Martin, and I. L. Humberstone. 1980. “Two Notions of Necessity.” *Philosophical Studies* 38 (1): 1–30. doi: 10.1007/bf00354523.
- Dudman, V. H. 1994. “Against the Indicative.” *Australasian Journal of Philosophy* 72 (1): 17–26. doi: 10.1080/00048409412345851.
- Edgington, Dorothy. 1995. “On Conditionals.” *Mind* 104 (414): 235–327. doi: 10.1093/mind/104.414.235.
- . 1996. “Lowe on Conditional Probability.” *Mind* 105 (420): 617–30. doi: 10.1093/mind/105.420.617.
- Gibbard, Allan. 1981. “Two Recent Theories of Conditionals.” In *Ifs*, edited by William Harper, Robert C. Stalnaker, and Glenn Pearce, 211–47. Dordrecht: Reidel.
- Grice, H. Paul. 1989. *Studies in the Way of Words*. Cambridge, MA.: Harvard University Press.
- Jackson, Frank. 1987. *Conditionals*. Blackwell: Oxford.
- . 1990. “Classifying Conditionals.” *Analysis* 50 (2): 134–47. doi: 10.1093/analys/50.2.134.
- . 1998. *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Clarendon Press: Oxford.
- Kripke, Saul. (1972) 1980. *Naming and Necessity*. Cambridge: Harvard University Press.
- Lewis, David. 1973. *Counterfactuals*. Oxford: Blackwell Publishers.
- . 1976. “Probabilities of Conditionals and Conditional Probabilities.” *Philosophical Review* 85 (3): 297–315. doi: 10.2307/2184045. Reprinted in his *Philosophical Papers*, Volume 2, Oxford: Oxford University Press, 1986, 133–152. References to reprint.
- . 1979a. “Counterfactual Dependence and Time’s Arrow.” *Noûs* 13 (4): 455–76. doi: 10.2307/2215339. Reprinted in his *Philosophical Papers*, Volume 2, Oxford: Oxford University Press, 1986, 32–52. References to reprint.
- . 1979b. “Scorekeeping in a Language Game.” *Journal of Philosophical Logic* 8 (1): 339–59. doi: 10.1007/bf00258436. Reprinted in his *Philosophical Papers*, Volume 1, Oxford: Oxford University Press, 1983, 233–249. References to reprint.
- . 1996. “Elusive Knowledge.” *Australasian Journal of Philosophy* 74 (4): 549–67. doi: 10.1080/00048409612347521. Reprinted in his *Papers in Metaphysics and Epistemology*, Cambridge: Cambridge University Press, 1999, 418–446. References to reprint.
- McCawley, James. 1996. “Conversational Scorekeeping and the Interpretation of Conditional Sentences.” In *Grammatical Constructions*, edited by Masayoshi Shi-

- batani and Sandra Thompson, 77–101. Oxford: Clarendon Press.
- Ramsey, Frank. 1929/1990. “Probability and Partial Belief.” In *Philosophical Papers*, edited by D. H. Mellor, 95–96. Cambridge University Press.
- Stalnaker, Robert. 1978. “Assertion.” *Syntax and Semantics* 9: 315–32.
- . 1981. “A Defence of Conditional Excluded Middle.” In *Ifs*, edited by William Harper, Robert C. Stalnaker, and Glenn Pearce, 87–104. Dordrecht: Reidel.

Published in *Philosophical Quarterly*, 2001, pp. 200-216.