ON THE TRUTH CONDITIONS OF INDICATIVE AND COUNTERFACTUAL CONDITIONALS

Wylie Breckenridge

In this essay I will survey some theories about the truth conditions of indicative and counterfactual conditionals. My aim is to give a taste of the issues involved, while pushing for no conclusion but this: they are difficult.

By an *indicative* conditional I mean any statement that can be expressed in the form "If A *is* the case then C *is* the case", and I will write these as A→C. By a *counterfactual* conditional I mean any statement that can be expressed in the form "If A *were* the case then C *would be* the case" and I will write these as A□→C. Three comments. First, indicative conditionals are written as A→C rather than as A⊃C to remove any presumption that they are material conditionals (even though they might turn out to be so). Second, *any* statement can be expressed in *either* of these forms if we are prepared to change its meaning, so I mean without changing its meaning or, if that seems problematic, at least in such a way that the person who made the statement agrees does not change its meaning. Third, "is the case" is intended to be tenseless - it should be understood as "is the case now, was the case in the past, and always will be the case in the future". Similarly for "were the case" and "would be the case". Specifications of time, where relevant, should be included in A and C.

Although it is usually clear whether or not a given statement is a conditional and, if so, what its antecedent and consequent are, it is by no means always clear whether it is intended to be an indicative or a counterfactual. I will not try to clarify the distinction any further here, but will do so when and only when needed throughout the essay.

PART I: INDICATIVE CONDITIONALS

A die is about to be rolled. I guess the result and make the following statement:

(1) If it's even then it will be a six.

Most of us would say that if a 6 is rolled then my statement was true, and that if a 2 or a 4 is rolled then my statement was false. But if a 1, 3 or, 5 is rolled then opinions tend to diverge. Some say that it was neither true nor false. Some say that it was true (recalling, perhaps, something from elementary logic). And some say that they cannot judge its truth or falsity until they know what the result *would* have been if it *had* been even. For present purposes we can ignore people in the last case. They seem to be interpreting my statement as the counterfactual conditional:

(2) If it were even then it would be a six,

and here we are interested in what people say about (1) when it is interpreted as an *indicative* conditional.

The example suggests the following: (i) we are prepared to judge indicative conditionals as true or false, so it makes sense to ask what their truth conditions are; (ii) we are prepared to make those judgments solely on the basis of the truth and falsity of the antecedent and consequent, so we should be able to give their truth conditions in terms of them alone; and (iii) indicative conditionals might sometimes lack truth value, even when their antecedents and consequents do not.

Mackie [1962] does not accept that indicative conditionals have truth conditions. He claims that they are expressions of condensed arguments rather than assertions of propositions, and explains observation (i) by saying that when we judge them true or false we are really judging whether or not they are *acceptable*. Suppose there is a cricket match this weekend. I know that the groundkeeper is proud of his pitch and that there is no alternative venue, so I assert the following:

(3) If it rains then the match will be cancelled.

I am, according to Mackie, expressing an argument which might more fully go something like this: "Suppose that it rains. Then if the match is played the pitch will be badly damaged. But the groundkeeper is proud of his pitch and will demand that the match is moved. But there is no alternative venue…", and so on. I am not asserting a proposition, not even the proposition that this argument exists. Rather, I am expressing the argument itself. If he is right then any search for *truth* conditions for indicative conditionals is in vain, and should, rather, be a search for *acceptability* conditions.

McDermott [1999, p. 295] asks how, if indicative conditionals are condensed arguments, can we explain their use as the antecedent or consequent of other indicative conditionals? Consider these compound indicative conditionals:

(4) If the match is cancelled if it rains, then the groundkeeper is proud of his pitch.
(5) If the groundkeeper is proud of his pitch, then if it rains the match will be cancelled.

McDermott claims that even though (5) may be thought of (in a somewhat ad hoc manner) as a condensed argument in which a second supposition is made within the scope of the first, no such treatment seems available for (4). My question to Mackie is this: what argument does he think I'm expressing when, in the case of the rolling die, I simply *guess* that if it's even then it will be a six? I can make the supposition that it's even alright, but I have no rational reason to go on to conclude that it will be a six. So any argument that I do offer ought to be judged unacceptable by any reasonable standard. So Mackie must say that if I make such a guess then it cannot be true. But that doesn't seem right. McDermott agrees with him that at least *some* indicative conditionals should be thought of as condensed arguments, but not all, and so we are left with the question of what, if any, are the truth conditions for the ones that aren't?

Adams [1975] agrees with Mackie that indicative conditionals do not have truth conditions, but for quite a different reason. As well as talking about the truth of statements it is common to talk about their *assertability* - the (subjective) extent to which speakers are prepared to assert (or assent to) them. It is widely accepted that for a non-conditional statement, A, its assertability, Ass(A), goes by its probability of truth, P(A). That is, a speaker is prepared to assert a non-conditional statement just to the extent that he believes the statement is *true*. Adams hypothesised (and it seems to now be widely accepted also) that for an indicative conditional statement, $A \rightarrow C$, its assertability, $Ass(A \rightarrow C)$, goes by $P(C/A)$, the conditional (subjective) probability of the consequent given the antecedent. But Lewis [1976] has shown that there is no reasonable way of assigning truth conditions to $A \rightarrow C$ such that $P(C/A) = P(A \rightarrow C)$ - that is, by Adams's hypothesis, such that $Ass(A \rightarrow C) = P(A \rightarrow C)$. For indicative conditionals, then, truth and assertability come apart. Adams concludes that since it is their assertability and not their truth that governs the way we use indicative conditionals, the important question is not when are they true, but when are they assertable? And that's a question he seems to have answered.

Lewis [1976, p. 85] admits that he has "no *conclusive* objection to the hypothesis that indicative conditionals are non-truth-valued sentences, governed by a special rule of assertability that does not involve their non-existent probabilities of truth" (my emphasis). He offers instead the *inconclusive* objection that the hypothesis would require too much of a "fresh start". It would require, for example, an account of the truth conditions of compound statements with non-truth-valued conditional components, or else an account of how all seeming examples of such compound statements can be explained away. With this is mind he suggests that we take indicative conditionals to have truth conditions, and to have the truth conditions that require the least amount of new work - those of the material conditional. The task for Lewis, then, is to explain why these truth conditions seem to give counter-intuitive results. If a five is rolled then my prediction (1) has a false antecedent and so, according to Lewis, it is true. So why do I not say so?

His answer, at least initially, is that what we are inclined to say is governed only in part by what we believe is true. There seems to also be, for example, a conversational maxim that we may call *assert the stronger*: if a speaker believes some statement Q to be true only because he believes another statement P to be to true then he should assert P (the stronger) instead of Q (the weaker). It is this, Lewis suggests, that accounts for the diminished assertability of some indicative conditionals. I am reluctant, he says, to assert that my prediction about the die was true, since I only believe it was true because I believe that it wasn't even, and I am conversationally obliged to assert that instead.

Let's assume that Lewis is right about indicative conditionals being material conditionals, so we can write them as $A \supset C$. How well does this agree with Adams's hypothesis (that $Ass(A \supset C) = Pr(C/A)$, or, consequently, that $A \supset C$ is assertable if and only if $Pr(C/A)$ is high)? Suppose I believe that $A \supset C$ is true, but only because I believe that $\neg A$ is true. That is, suppose that for me $P(A \supset C)$ is high, $P(\neg A)$ is high and $P(A \supset C/A)$ is low. According to Lewis, $A \supset C$ is not assertable because $\neg A$ is stronger than $A \supset C$ and should be asserted instead. According to Adams's hypothesis, $A \supset C$ is not assertable because $P(C/A) = P(A \supset C/A)$, which is low. So the two agree in this case. Now suppose I believe that $A \supset C$ is true, but only because I believe that C is true. That is, suppose that for me $P(A \supset C)$ is high, $P(C)$ is high and $P(A \supset C / \neg C)$ is low. According to Lewis, $A \supset C$ is not assertable because C is stronger than $A \supset C$ and should be asserted instead. According to Adams's hypothesis, if $P(C/A)$ is high then $A \supset C$ is assertable. But it *is* possible in this case for $P(C/A)$ to be high. Suppose that I plan to go to the movies tonight whether or not it rains. Let A be "it will rain" and C be "I will go to the movies". Then $P(C)$ is high (I believe C), $P(A \supset C)$ is high (I believe $A \supset C$ because I believe C), and $P(A \supset C / \neg C)$ is low (I believe $A \supset C$ *only* because I believe C). But

P(C/A) is high, because I believe that I will go to the movies even if it rains. Bad news for Lewis: he disagrees with Adams's hypothesis here.

Lewis concedes the point, suggesting as an explanation that "considerations of conversational pointlessness are not decisive … creating only tendencies towards diminished assertability … that may or may not be conventionally reinforced" (p. 88). But in a postscript to his original article, Lewis retracts this theory in favour of an alternative advanced by Frank Jackson.

Jackson [1979] agrees with Lewis that the indicative conditional has the truth conditions of the material conditional, but suggests that their diminished assertability is a result of a particular *conventional* use of the conditional construction (rather than a general conversational maxim). The problem for Lewis's theory is that it says that $A \supset C$ has diminished assertability in certain cases not because of any particular feature of the conditional construction, but because of a feature of conversation in general. It thus allows the antecedent and consequent to have an equal effect on the assertability of the conditional. But if Adams's hypothesis is right then they don't, as was demonstrated in the previous paragraph. Jackson's suggestion is that the conditional construction $A \rightarrow C$ is used in conversation to express the truth of the material conditional $A \supset C$, and also, by convention, to signal that $A \supset C$ is *robust* with respect to A, by which he means that $P(A \supset C/A)$ is high. So according to Jackson, $A \rightarrow C$ is assertable if and only if $P(A \supset C)$ is high and $P(A \supset C/A)$ is high. But these are both high if and only if P(C/A) is high. So the theory says that $A \rightarrow C$ is assertable if and only if P(C/A) is high, in exact accordance with Adams's hypothesis. Jackson's advantage over Lewis here comes from taking the assertion of a conditional as signaling *only* that $P(A \supset C/A)$ is high (and not signaling anything further about $P(A \supset C/\neg C)$ - that is the task, he says, of constructions like "A or anyway C").

McDermott [1996, pp. 20-23] has a number of objections to Jackson's theory, but one of them seems to pose a particularly serious problem. Even though the way we *assent* to indicative conditionals seems to support the theory, the way that we make judgments about *betting* does not. Suppose, in the case of the die, that by saying (1) I am placing a bet, and that it turns out to be a five. It seems reasonable to expect that whether or not I win the bet depends not upon the *assertability* of my prediction but upon its *truth*. If that is right then according to Jackson it should be agreed all round that my prediction was true and that I win the bet. But most people say that I do *not* win the bet - that it is, in fact, called off. Why would they say that if they think my prediction was true? Jackson's reply seems to be that our use of indicative conditionals is governed only by judgments of assertability and not at all by judgments of truth, and that truth conditions are needed only to explain the rule of assertability. But to that McDermott asks why do we need to have truth conditions whose *only* use is to explain the rule of assertability? Why not just have the rule of assertability by itself? Jackson cannot reply that by dropping truth conditions for indicative conditionals we would be driving a wedge between them and non-conditionals, because his own theory already does that - by claiming that the use of conditionals, unlike non-conditionals, is governed by a special rule of assertability and not by truth conditions.

McDermott's own theory [1996] is that $A \rightarrow C$ has the same truth value as C if A is true, and lacks truth value otherwise. Thus he agrees with Jackson that the indicative conditional has truth conditions, but he disagrees about what they are. In particular, he allows that the conditional be non-truth-valued in some cases. This theory has great explanatory advantages over Jackson's. If we take the probability of a statement (conditional or not) to be the probability that it is true, and take the assertability of a statement (conditional or not) to be the probability that it is true given that it has truth value, then we get Ass(A) = P(A) as usual for statements that cannot lack truth value, and we get $\mathrm{Ass}(A \rightarrow C) = P(A \wedge C/A) = P(C/A)$. Thus McDermott can *explain* (not just agree with) Adams's hypothesis, and also why it is that probability and assertability come apart in the case of indicative conditionals. The theory also shows that Adams was too hasty to conclude that any theory about the truth conditions for $A \rightarrow C$ such that $\mathrm{Ass}(A \rightarrow C)$ is not $P(A \rightarrow C)$ is unsatisfactory. Furthermore, whereas Jackson needs to make some rather unappealing claims to account for our judgments about betting on conditionals, this theory not only agrees with but explains them.

How well the various theories can be made to work is not, however, the only thing that might be considered by anyone interested in choosing between them. Those of Lewis and Jackson simplify the logic of indicative conditionals (and, therefore, of logical connectives in general), but they complicate the explanation of verbal behaviour. That of McDermott's complicates the logic (by requiring, to start with, an extension of the truth tables for the other logical connectives to allow for truth value gaps), but it simplifies the explanation of verbal behaviour. Considerations of simplicity may be involved as well.

PART II: COUNTERFACTUAL CONDITIONALS

A die has the numbers 1, 3, 5, 6, 6, and 6 on its faces. It is rolled and comes up 5. Consider the following two counterfactual statements:

(6) If it had been even then it would have been a six.
(7) If it had been even then it would have been a four.

Most of us would agree that (6) is true and (7) is false. So it seems that counterfactuals, too, have truth conditions - conditions that are satisfied by (6) but not by (7). But in both of these the antecedent and consequent are false (it wasn't even, it wasn't a six, and it wasn't a four). So, unlike indicative conditionals, the truth conditions must involve *more* than just the truth values of the antecedent and consequent.

It seems to be a matter of logic that (6) is true and (7) is false: if the number had been even then it is logically impossible for it to have been anything but a six. But not all true counterfactuals can be accounted for in this way. Suppose that a dry, well-made match sat on my desk in the presence of oxygen all day yesterday without being struck. Then I would say that the following counterfactual is true:

(8) If the match had been struck then it would have lit.

Why is that true? Easy. Because if the match had been struck it would still have been dry, and if the match had been struck it would still have been well-made, and if the match had been struck it would still have been in the presence of oxygen, and all dry, well-made matches in the presence of oxygen light. But this answer appeals to the truth of three counterfactuals. Why are *they* true? Easy. If the match had been struck it would have still been dry, because no dry match has to become wet in order to be struck. If the match had been struck it would have still been well-made, because no well-made match has to become not-well-made in order to be struck. If the match had been struck it would have still been in the presence of oxygen, because no match in the presence of oxygen has to cease to be in the presence of oxygen in order to be struck. No counterfactuals there - just an appeal to the truth of four laws about the behaviour of matches. But by appealing to these laws I have actually taken a further four counterfactuals to be true: if the match had been struck then it would still have been true that no dry match has to become wet in order to be struck, and if the match had been struck then it would still have been true that no well-made match has to become not-well-made in order to be struck, etc. But aren't these counterfactuals automatically true because laws are supposed to hold universally? No. Sometimes laws get dropped in counterfactual situations, such as when we decide whether or not it is true that if the gravitational constant had been slightly different then intelligent life would not have evolved on Earth. So the answer still contains counterfactuals whose truth remains to be explained.

In general, it is surprisingly difficult to say what makes a counterfactual true without appealing to the truth of other counterfactuals - to give an account of their truth conditions, that is, which is *non-circular*. (There is a further problem with the attempted explanation: even if it *is* true, for example, that no dry match has to become wet in order to be struck, why is it true that if the match had been struck it wouldn't have just *happened* to become wet? Even if we can remove the circularity, we still need to answer that question as well.)

Just as he does for indicative conditionals, Mackie [1962] denies that truth conditions can be given by claiming that counterfactual conditionals are just condensed arguments, so that we should be asking for their acceptability conditions instead. Moreover, he claims that he can give these conditions in a way that is non-circular. But McDermott [1999, p. 295] argues against this in the same way that he does for indicative conditionals. He asks how, if counterfactual conditionals are condensed arguments, can we explain their use as the antecedent or consequent of compound conditionals? Consider:

(9) If the match would have lit if it had been struck, then oxygen was present.
(10) If oxygen was present, then the match would have lit if it had been struck.

Again, even though (10) can possibly be thought of as a condensed argument in which a second supposition is made within the scope of the first, no such treatment seems available for (9). And I repeat my question to Mackie: what if I knew nothing about the die and simply *guessed* that if it had been even it would have been a six? Surely the fact that I have made a guess makes it no less true? But I could have *no* acceptable argument from the supposition that the result was even to the conclusion that it was a six.

As for indicatives, McDermott (pp. 296-299) agrees with Mackie that at least *some* counterfactual conditionals should be thought of as condensed arguments, but not all, and so we are left with the question of what, if any, are the truth conditions for the ones that aren't? (He also argues (pp. 292-294) that Mackey fails, anyway, to give a non-circular account of their acceptability conditions.)

Lewis [1973; 1986, Ch. 17, Postscripts] believes that counterfactuals have truth conditions. He says that A□→C is vacuously true (at our world, w_0) if and only if there is no possible world in which A is true, and non-vacuously true if and only if there is a possible world w_1 at which A and C are both true, and which is more similar to w_0 than any other world w_2 at which A is true and C is false (That is, if and only if every closest A-world is a C-world). He says that our judgments about the similarity of possible worlds seem to be constrained by the following:

> (L1) It is of the first importance to avoid big, widespread, diverse violations of law.
> (L2) It is of the second importance to maximize the spatio-temporal region throughout which perfect match of particular fact prevails.
> (L3) It is of the third importance to avoid even small, localized, simple violations of law.
> (L4) It is of little or no importance to secure approximate similarity of particular fact, even in matters that concern us greatly. (1986, p.48; labeling altered to follow McDermott)

His theory seems to assign the correct truth value to (8): Of all the worlds in which the match was struck, the ones most similar to the actual world (according to L1 - L4) are those in which (a) the match was still dry when struck, (b) the match was still well-made when struck, (c) the match was still in the presence of oxygen when struck, and (d) all dry, well-made matches in the presence of oxygen light when struck, and in all of these the worlds the match lights when struck.

Most of us evaluate counterfactuals under the assumption that things could have gone differently, and could have done so without miracles. That is, we evaluate counterfactuals under the assumption that the laws of nature are *indeterministic*. So truth conditions that give intuitively wrong verdicts when indeterminism is assumed cannot be the ones that we actually use. A good theory of the truth conditions of counterfactuals must work in an indeterministic world.

Lewis's theory, it seems, does not. McDermott illustrates this via two examples. The first [1999, p. 304] he attributes to Morgenbesser. An (indeterministic) coin is about to be tossed. I bet on heads, it lands tails, and I lose. Then the following is intuitively true:

> (11) If I had bet on tails then I would have won.

But Lewis's theory says that this is false. Why? Of all the closest worlds (according to L1-L4) in which I bet on tails, L1-L4 cannot distinguish between those in which the coin still lands tails (so that I win) and those in which the coin lands heads (so that I still lose). Why? First, the coin toss is indeterministic so no law is violated whether the coin lands heads or tails. So they cannot be distinguished by L1 or L3. Second, in both types of world the region of spatio-temporal perfect match with the actual world ends at the moment of the bet, so they cannot be distinguished by L2. Third, if it is of *no* importance to secure approximate similarity of particular fact then we cannot use L4 to distinguish them. If it is of *some* importance, then those worlds in which the coin lands tails agree with the actual world in one matter of fact (the result of the toss) and disagree in two matters of fact (the bet placed and the result of the bet), but those worlds in which the coin lands heads likewise agree with the actual world in one matter of fact (the result of the bet) and disagree in two matters of fact (the bet placed and the result of the toss). So even then L4 cannot distinguish them. So it is not true in all of the closest worlds in which I bet tails that I win, and the counterfactual is false.

The problem for Lewis here is that he needs to keep his theory *non-circular*. Intuition says that what happens in the spatio-temporal neighbourhood of the coin toss is *independent* of what happens in the spatio-temporal neighbourhood of the bet, and that if I had bet tails then the coin would still have landed tails. It would help Lewis's theory if he could allow that worlds be compared for similarity on a region-by-region basis rather than as spatio-temporal wholes, where the regions to be compared are independent (in some relevant respects). But there seems to be no way of specifying what it is for two regions to be independent without appealing to counterfactuals (such as what would (or would not) happen in one region as the result of what happens (or does not happen) in the other).

The second case that McDermott cites shows, he suggests, that Lewis faces quite a different problem. X atoms have a chance of 1 in a million of decaying spontaneously in the next minute. The chance of decay is doubled if the atom is subject to radiation. On this occasion the atom is subject to radiation and it does decay within a minute. Consider:

(12) If radiation had not been present then the atom would not have decayed.
(13) If radiation had not been present then the atom would (still) have decayed.

According to McDermott, intuition says that one of (12) and (13) is true, but we don't know which. What does Lewis say? Consider our closest worlds (according to L1-L4) in which radiation was not present. As for the Morgenbesser example, L1-L3 cannot distinguish between those in which the atom decays and those in which it does not. But L4 might. If approximate match is of any importance, then the worlds in which the atom decays are more similar to the actual world than those in which it does not. In that case (12) is false and (13) is true. This agrees with intuition in saying that exactly one of them is true, but it disagrees in giving a definite verdict. If approximate match is of no importance then worlds in which the atom does not decay are just as similar to the actual world as those in which it does. In that case (12) and (13) are both false. This disagrees with intuition both in saying that neither of them are true, and again in giving a definite verdict. Either way, Lewis is in trouble.

McDermott thinks that if he is to avoid counterintuitive results in other situations then Lewis has to take approximate similarity to be of no importance, so that (12) and (13) are both false. But to say that they are both false is to say that there is no fact of the matter about what would have happened if the radiation had not been present - it is to say that there is no counterfact. Lewis cannot avoid this conclusion. His theory is reductivist - it says that counterfactuals are made true by nothing more than facts about the actual world[1]. In the case of the X atom, there are no facts about the actual world (including its laws) that say what would have happened if the radiation had not been present. And for Lewis, no fact of the matter in the actual world is no fact of the matter simpliciter. But McDermott thinks this is counterintuitive - that we tend to think that there is fact of the matter about what would have happened.

McDermott [1999] concludes that Lewis's reductivist theory cannot work and proposes his own, realist, theory. It is based on the primitive notion of an *access point* at a world - a point at which things can go (could have gone) in different ways. Moreover (and this is what makes it a realist theory), at each point one of those different ways is singled out as the way that things *will* go (*would have* gone) - there is a fact of the matter at every access point in every possible world. Counterfactual statements are claims about these facts. Just as, according to some theories, the statement "the cat is on the mat" is made true by the fact that the cat *is* on the mat, the statement "if the match had been struck it would have lit" is made true, according to McDermott, by the fact that the match *would* have lit if it had been struck. The theory promises an easy account of the logic of counterfactuals and of our verbal behaviour. But it carries a metaphysical cost - a commitment to facts over and above those about the actual world. Lewis has to work harder to get his theory to work, if it can be made to work at all. But it is metaphysically cheaper - it needs no extra facts. For indicative conditionals we saw that considerations of simplicity may play a role in choosing between theories about their truth conditions. For counterfactuals, metaphysics may play a role as well.

REFERENCES

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

GOODMAN, N. (1979), *Fact, Fiction and Forecast*.

JACKSON, F. (1977), "A Causal Theory of Counterfactuals", *Australasian Journal of Philosophy*, **55**, pp. 3-21.

—— (1979), "On Assertion and Indicative Conditionals", *Philosophical Review*, **88**, pp. 565-89; reprinted in F. Jackson (ed.), *Conditionals*.

LEWIS, D. (1973), *Counterfactuals*, Oxford: Blackwell.

—— (1976), "Probabilities of Conditionals and Conditional Probabilities", *Philosophical Review*, **85**, pp. 297-315; reprinted in F. Jackson (ed.), *Conditionals*.

—— (1979), "Counterfactual Dependence and Time's Arrow", *Nous*; reprinted in *Philosophical Papers, v. II* (Oxford: Oxford University Press, 1986).

[1] Contrary to this, his theory seems to say that what makes a given counterfactual true are facts about our most similar antecedent worlds. But Lewis would say that these facts are determined, in turn, by facts about the actual world - any world exactly like ours in matters of fact has most similar antecedent worlds that are exactly like our most similar antecedent worlds in matters of fact. (This is roughly, but not quite, his view: Lewis thinks, in fact, that no two distinct worlds can be exactly alike.)

MACKIE, J. L. (1962), "Counterfactuals and Causal Laws", in R. J. Butler (ed.) *Analytical Philosophy*, Oxford: Blackwell, pp. 66-80.

MCDERMOTT, M. (1996), "On the Truth Conditions of Certain 'If'-Sentences", *The Philosophical Review*, **105**, pp. 1-37.

—— (1999), "Counterfactuals and Access Points", *Mind*, **108**, pp. 291-334.