Conditionals and Relativism

- 1. Initial Arguments for Epistemic Theories
- 2. Inference Patterns and Epistemic Explanations
- 3. Three Types of Epistemic Theory
- 4. Sly Pete and Direct Evidence for Relativist Account
- 5. Disagreement

Draft version 30 June 2005 – References etc incomplete Brian Weatherson, Cornell University

I'm going to defend a relativist account of the semantics of indicative conditionals. The main argument of the paper has the following form.

- 1. An epistemic theory of the truth conditions of indicative conditionals is correct.
- 2. If an epistemic theory of the truth conditions of indicative conditionals is correct, then the correct semantic theory for indicatives is relativist.
- 3. So, the correct semantic theory for indicatives is relativist.

Since premise 1 has been defended at length elsewhere (Stalnaker 1975, Davis 1979, Weatherson 2001, Nolan 2003, Chalmers ms) I won't say *very much* about it here. In section 1 I will say more precisely what I take an epistemic theory to be, and what I take the strongest arguments for it are. The main point of the paper is to defend premise 2.

1. Initial Arguments for Epistemic Theories

It is traditional to start a paper on indicative conditionals with a ritual denunciation of theories that analyse the indicative as a material implication, and I'm going to respect tradition here. (Everything in this section should basically be reminding people of stuff they already know – if I were arguing for theses from scratch here I'd move at a much slower pace.) Consider the following little exchange, which we imagine taking place as I'm giving a talk.

Brian: The Pope is here.

Audience member: If what Brian just said is true, the Pope is not here.

Intuitively, what the audience member said is *false*. But if the indicative conditional *If A, B* is true just in case *A* is false or *B* is true, then what the audience member said is *true*, since it is an indicative conditional with a false antecedent. The best response that I know of to this objection is due to Frank Jackson (1987). Jackson argues that in these cases what the audience member says is *true*, but it isn't *assertible*¹. The reason for the latter is that for an indicative conditional to be assertible, the probability of its consequent given its antecedent has to be high. This last claim seems to be too strong, and hence not explanatory, because of inferences like the following.

(4) Sally earns twice as much per hour as Suzy. So if Sally earns \$8 per hour, then Suzy earns \$4 per hour.

1

¹ This is a technical term of Jackson's. A sentence is *assertible* iff it is *assertable*, or it fails to be *assertable* for non-semantic reasons. The latter clause is there so that ordinary sentences that express truths but which it would normally be pointless or inappropriate to assert as still assertible. So a sentence like "Sydney is north of Melbourne" is still assertible, even in the context of a conversation about epistemology (to which it is irrelevant) or in a silent reading room (where nothing should be asserted).

Whether or not we think this inference is *truth* preserving, it definitely seems to preserve *assertibility*. Whenever we are in a position to assert the premise, we are in a position to assert the conclusion. But it's possible that the premise could be known to be true (and hence easily assertible) even though it is also held to be *extremely* probable that everyone earns over \$6 per hour (say because it's the local minimum wage) and hence that it is very improbable that Suzy earns \$4 per hour given that Sally earns \$8 per hour. So the best explanation of what's wrong with our audience member's assertion (assuming it is true) fails, and the natural conclusion to draw is that the assertion is actually false. Hence it follows that indicative conditionals are not equivalent to the corresponding material implications.

A natural thought at this point is to move to some kind of possible worlds account of indicatives, similar to the account offered by Stalnaker (1968) and Lewis (1973) for subjunctives. The problem for this, as Jackson (1987) noted, is that some indicatives seem to resist any possible worlds analysis. Consider the following conspiracy theory. It turns out the stuff in the rivers, lakes and oceans, the stuff to which we normally apply the term 'water', really is atomic after all. The theory that it is molecular, that it is H₂O, is a rumour spread by scientists who wanted to show that they were ever so smarter than the ancient Greeks. Call this story The Conspiracy Story, and consider (5).

(5) If The Conspiracy Story is true, then water is atomic.

It seems (5) is true. But on a possible worlds account of indicative conditionals, (5) should be extremely problematic. For the antecedent of (5) is true at some possible worlds, but the consequent of (5) is false at every possible world. So by *any* measure of similarity, it isn't true that at the nearest worlds where the antecedent of (5) is true, the consequent of (5) is true.

One natural move at this point is an epistemic account of indicatives.² Even if it is necessarily false that water is atomic, it isn't knowable *a priori* that water is atomic, so there are 'epistemic possibilities' (i.e. sets of propositions that we don't know *a priori* aren't all true) in which water is atomic. So its natural given the truth of (5) to think that the role played by possible worlds in the Stalnaker-Lewis theory of subjunctives should be played in a theory of indicatives by epistemic possibilities.

Epistemic theories of indicatives have the following clause as a non-trivial constituent.

(6) If A, B is (determinately) true if $\neg (A \land \neg B)$ is better supported by the evidence than $\neg (A \land B)$

This is a schema, which needs to be filled in at many points. In particular it leaves open the following four questions.

- Are there any other clauses in the definition? A purely epistemic theory would strengthen the 'if' used in the definition to an 'iff'. I prefer a theory in which $A \wedge B$ entails If A, B, and $A \wedge \neg B$ entails $Not \ if A$, B, so this epistemic clause only applies when A is false. (See Weatherson 2001 for more details.) But I regard theories on which (6) plays a non-trivial role as epistemic, even if they are supplemented by principles like this.
- When is *If A, B* false? Some theorists (e.g. Nolan 2003) argue that we should accept bivalence here, so *If A, B* is false iff it is not true. I prefer to follow Stalnaker in saying that at least for epistemically possible *A, If A, B* is false iff *If A, not B* is true, so in many cases *If A, B* will have an indeterminate truth value.

2

² As Chalmers, Nolan and Weatherson have noted, one can use the resources of two-dimensional modal logic to frame such a theory as a non-standard possible-worlds story, so the contrast with the previous paragraph is perhaps a little forced.

- What does 'better supported by the evidence' mean in this context? My preferred way of cashing this out is that p is better supported than q iff the agent with this evidence is in a position to know p but not in a position to know q. Others may prefer a more fine-grained account of support, perhaps to account for intuitions that some instances of strengthening the antecedent are invalid.
- Whose evidence matters here?

Although the first three questions are fascinating, I'm going to completely bracket them here so as to concentrate on the fourth. For it is on the fourth question that we get an argument for relativism. Before we get there, we need to look at another source of evidence that looks *prima facie* like a problem for an epistemic theory, but which on closer inspection turns out to support a relativist epistemic theory.

2. Inference Patterns and Epistemic Explanations

Each of the following inferences seems to be perfectly reasonable, in that a speaker who reasonably accepts the premises seems entitled to the conclusion.

- (7) Either the butler did it or the gardener did it. So if the butler didn't do it, the gardener did.
- (8) All toves are mimsy. So if Papi is a tove, Papi is mimsy.
- (9) Jack and Jill have the same favourite team. So if Jack's favourite team is the Cyclones, Jill's favourite team is the Cyclones.
- (10) Peter and Paul have different mothers. So if Paul's mother is Mary, Peter's mother is not Mary.

Here are the patterns that these inferences are instances of.

- (7a) A or B So if not A, B
- (8a) All Fs are Gs So if a is an F, a is a G
- (9a) f(a) = f(b)So if f(a) = x, f(b) = x
- (10a) $f(a) \neq f(b)$ So if f(b) = x, $f(a) \neq x$

The existence of plausible patterns like this is often taken to be evidence that the indicative is true iff the corresponding material implication is true. To see this, assume (a) that it takes more than the truth of the corresponding material implication for the conditional to be true, but (b) these inferences preserve truth. Then whatever 'extra' it takes for the conditional to be true must also be required for the premise to be true. But most of the standard candidates for the extra condition are not guaranteed by the premises. This is most obvious in the case of (7), but in fact it seems to be true for most of the inferences in question. In all four cases, neither the truth nor the high probability of the premises entails, for instance, the high probability of the consequent given the antecedent. The last two inferences, taken together, entail that for any f, a, b and x, the disjunction (11) has to be true.

(11)
$$(f(a) = x \rightarrow f(b) = x) \lor (f(b) = x \rightarrow \neg f(a) = x)$$

If we let f be a function from propositions to their truth values, let a and b be propositional variables, let x be truth, and assume that f(p) = true is equivalent to p (as seems safe), then as an instance of (11) we get (12).

$$(12) \qquad (p \to q) \lor (q \to \neg p)$$

Even if p and q concern completely separate subject matters, each have probability one-half and are probabilistically independent, one of these two has to be true, assuming that inference patterns (9a) and (10a) are always truth preserving. That is to say, if (12) is always true, then there is no plausible 'extra' property that a conditional has to have in order to be true beyond the truth of the corresponding material implication. That's to say the only plausible theory of conditionals on which (12) is a logical truth is the theory that indicative conditionals are true iff the corresponding material implication is true.

But these are not the only kinds of implications concerning conditionals that are intuitively plausible. There are also implications where we properly conclude that certain conditionals are false. Richard Bradley (2000) discusses the following case.

[O]ne cannot be certain that B is not the case if one thinks that it is possible that if A then B, unless one rules out the possibility of A as well. You cannot, for instance, hold that we might go to the beach, but that we certainly won't go swimming and at the same time consider it possible that if we go to the beach we will go swimming! To do so would reveal a misunderstanding of the indicative conditional (or just plain inconsistency). (2000: 220)

More generally, someone who regards A as an epistemic possibility, but knows that B is false, should regard If A, B also as something they know to be false. Bradley puts this in probabilistic terms as follows.

Preservation Condition: If
$$Pr(A) > 0$$
 but $Pr(B) = 0$, then $Pr(A \rightarrow B) = 0$

This isn't obviously the best formulation of his principle. In the example, what matters is not that A has non-zero probability, but that it is something that might be true. (One might try arguing these are equivalent, but I think that would be an uphill battle, in part because of all sorts of this that have zero probability that might be true.) The structure of the implication looks to be what is given in (13), where Kp means the relevant agents knows that p.

(13)
$$\neg K \neg A$$

 $K \neg B$
So, $K \neg (A \rightarrow B)$

But this is not valid if the conditional is a material implication. So now it is *impossible* to accept all of the intuitively plausible principles about implications involving conditionals are truth-preserving. There must be some other explanation of the reasonableness of all these implications.

The best explanation I know of is the one endorsed by Daniel Nolan (2003) as an explanation of implications like (7). Nolan says that given an epistemic theory of the indicative, we can say that each of the implications has the following property. Any speaker who knows the premise is in a position to truly assert the conclusion. Call an implication like this, where knowledge of the premise implies truth of the conclusion, epistemically acceptable. If we are confusing valid implications with epistemically acceptable ones, this could explain why all of (7) through (10) seem reasonable. More impressively, this hypothesis of Nolan's explains why (13)

seems reasonable, given an epistemic theory of indicatives. If we know that A is true in some epistemic possibilities, but B is false in all of them, then we are in a position to know that $\neg(A \land B)$, but $A \land \neg B$ is consistent with the evidence. So it's definitely *not* the case that $\neg(A \land \neg B)$ is better supported than $\neg(A \land B)$. (I assume here that the epistemic clause is the salient clause for working out the truth value of If A, B. It's obviously hard to say something that covers all possible theories, but it is hard to see how any other clause of a broadly epistemic definition could make the conditional true in such a case.) So (13), like (7) through (10), is epistemically acceptable.

3. Three Types of Epistemic Theory

Let's come back to the hard fourth question we asked above. If the truth of an indicative conditional is determined by the relative evidential support of $A \wedge B$ and $A \wedge \neg B$, then we need to find out what the evidence is in this equation. Since the evidence in question is presumably someone's, or some group's, we'll need to identify the person or group. There are three families of solutions here: invariantist, indexicalist, and relativist.

The invariantist solutions say that there is one privileged group, and their evidence is always what matters for determining whether a particular utterance of If A, B is true. That is, there is some group G such that If A, B is true (in the relevant cases) if $\neg (A \land \neg B)$ is better supported by the evidence available to any member of G than $\neg (A \land B)$.

If we want to keep Nolan's explanation of the inferences in the previous paragraph, then G must include all speakers. For any speaker can infer that If A, B is true given that they know something that, when combined with A, entails that B. If that speaker is not part of G, they could not make this inference. But now we're back very close to the material implication theory. Indeed, if we assume that G includes G od, who knows that N or B is true iff it is true, then this theory implies that the indicative is true iff the corresponding material implication is true. Even without theism, this position looks too strong. When I deny that a conditional like (14) is true.

(14) If the American President is a spy, he's working for the Australian government

I certainly don't mean to be denying that there is anyone in the world who knows the disjunction, *Either the American President is not a spy, or he's working for the Australian government.* So I don't think there's much chance of an invariantist epistemic theory being plausible.

In my earlier work, I assumed that some kind of indexicalist theory is correct. An indexical theory says that there is some function f from speakers to evidence such that an utterance by S of If A, B is true (in the relevant cases) if $\neg (A \land \neg B)$ is better supported by the evidence available to f(S) than $\neg (A \land B)$. We can generate different kinds of indexical theories varying f. One indexicalist theory takes f to be the identity function, so the truth of the conditional depends just on the speakers' evidence. Another theory takes it to be the speaker and her intended audience, another the speaker and the people she is talking about, and yet another the speaker and everyone who hears the utterance. By looking at how we evaluate utterances by other people, we will be able to see which of these theories is most plausible.

As several of the examples Nolan provides show, the theory that f is the identity function is not very plausible. Here's one of his examples.

I know our glorious leader has troops especially fearsome in the charge, and which, with her leadership, are more than a match for the troops of the enemy. I am awaiting news of the battle, but I confidently predict "if she charged, she won". I then receive late spy reports of the enemy's deposition, which report that the enemy have chosen the battlefield carefully so that a charge would be

disastrous (it may be that there is a marsh in the middle of the field which troops will not notice until they stumble into it, or that there are concealed pits and furrows, or perhaps the field is full of carefully concealed mines). I might then not only then be prepared to claim "if she charged, she lost", but be prepared to say that my previous claim was false (and I might hope that my leader did not share my belief). It seems plausible that it would be right of me to take myself to have said something false, not merely that I said something which, if said now, would express a falsehood. (Nolan 2003: 229)

If f were the identity function, then when I learn more information, I should now think that my earlier prediction was true. For on that view, my earlier prediction was a claim much like (15).

(15) I have good all-things-considered evidence that she did not charge and lose.

And I can think that an earlier utterance of (15) is true, even if I now get more evidence that undermines my previous confidence.

We can generalise Nolan's story. Imagine the person who finds out that the battlefield is unfriendly is not a later stage of myself, but a third party who hears my comments. She can still say that my prediction is false. Indeed, even if she hears not my comments, but a recording of them, she can still say that my prediction is false if she has good evidence that the leader did not charge and win, and little evidence that the leader did not charge and lose.

Nolan suggests that the explanation for this is that the epistemic theory has to be supplemented by some "objective component to [the] evaluation of conditionals" (229). This seems like the wrong reaction to me. Imagine the person hearing a recording of my words does not know about the state of the battlefield, but is an historian writing a biography of that day's enemy general, a figure famous for never losing a battle in his long career. The historian, if she does not know whether a battle took place that day or not (imagine the historian has dated my recording to the time our fearless leader was frequently matched up against the enemy general, but does not know just which day it was) will still deny that my prediction was true. But in this case there need be no objective features that undercut my claim. It may well have been that had the glorious leader charged she would have won, but she failed to do so, and hence the enemy general's record was kept intact.

The conclusion we seem pushed towards is that *anyone* who hears my utterance of a conditional can use her evidence to evaluate what I say. This suggests that f is a function from the speaker to whoever is in her audience. (This is what I more or less accepted in the previous paper.) But this is a rather incredible position. If it is true, then the truth value of any indicative conditional in the paper (including this one!) is determined in part by whether anyone is using their iPod to record my words, and whether those words will be played back for future generations to hear and evaluate. It seems highly implausible that the truth of my words could depend on the future in quite *this* way. We shall return to this view later to try and strengthen this argument from implausibility, but first I want to turn to the relativist option.

The relativist says that the truth of a particular conditional is relative to an evaluator. In particular, in cases where the epistemic clause of the truth conditions is relevant, the relevant evidence is that of the evaluator. So if S says If A, B and B hears this, then (in relevant cases) B should judge that as true if $A \cap B$ is better supported by her evidence than $A \cap B$. And by 'should' here I mean that B judges the truth of the conditional (relative to her situation) correctly if she compares the status of those two propositions with respect to her evidence.

6

³ There is a more sophisticated relativist theory where the relevant evidence is not the evaluator's, but the evidence of some group that is the value of a function with the evaluator as argument. So it may be, for example, the evaluator and all those in her community, or the evaluator and all those she has easy access to,

This explains nicely what is going on in Nolan's example. When I learn that the battlefield was not set up to favour the glorious leader, I move into a new context relative to which my old prediction is now incorrect. Similarly the historian who knows the enemy general never lost a battle is in a context relative to which my old prediction is incorrect. But this is not to say I made any kind of mistake in making the prediction. It may even be that relative to my original context, my prediction was *true*, because the truth of the prediction is relative to a context of evaluation.⁴

The relativist also has a nice generalisation of Nolan's explanation of the inference patterns discussed in section 2. It seems in each case that *anyone* who knows the premise can properly conclude that the conclusion is true, even if they aren't the one who uttered the conclusion. If I know that Jack and Jill were born in the same country, I can conclude that your utterance *If Jack was born in Canada, so was Jill* is true. That's because, says the relativist, if I know that Jack and Jill were born in the same country, then your utterance is true relative to my context.

4. Sly Pete and Direct Evidence for Relativism

One of the central difficulties in formulating a theory of conditionals is explaining the intuitions about cases like Allan Gibbard's "Sly Pete" example (Gibbard 1981: 231-2). Here is Jonathan Bennett's version of the case, slightly modified to remove some potential distractions.

A hand of poker is being played, and everyone but Pete and one other player have folded. Two onlookers leave the room a this point, and a few moments later each sees one player leave the gaming room. Winifred sees Pete without the scowl and trembling cheek that he always has after calling and losing, and concludes that *If Pete called, he won*; Lora sees Pete's opponent caressing more money than he owned when Lora left the room, and concludes that *If Pete called, he lost*. (Bennett 2001: 83)

Bennett takes this case to be evidence for his theory that indicative conditionals lack truth values. The argument for this conclusion is that any assignment of truth values to the conditionals seems to be implausible in one way or another. But Bennett only considers non-relativist theories; a theory on which indicatives have truth values relative to a context of evaluation is immune to the argument that he offers.

There are two data points that a theory of conditionals has to take into account here. First, Winifred and Lora seem to be *disagreeing* when they utter these conditionals. They are certainly inclined to *deny* the other person's conditional. If you ask Winifred, is it the case that if Pete called he lost, she will say no. Second, both Winifred and Lora make no mistakes. They make what seems to be a perfectly reasonable inference from correct observations. Pete does not have the telltale scowl because he didn't call and lose. And Pete's opponent seems to have more money because he really has more money. (In case it isn't clear, Pete really did fold in the story.)

But it seems that it will be hard to account for both of these intuitions in a theory where conditionals have truth values. If conditionals are apt to be true, and Winifred makes what looks like a flawless inference to a conditional from true premises, then the conditional she comes to believe is true. By similar reasoning, Lora's conditional is true. But if two people each have true beliefs, they cannot be thereby disagreeing. The only way out of this trap, suggests Bennett, is to give up on the idea that conditionals are truth valued.

or some other group centered on the evaluator. Such complications will be set aside for the purposes of this paper.

For more general defenses of the plausibility of this kind of relativist theory, see MacFarlane (xxxx, xxxx, xxxx), and Egan, Hawthorne and Weatherson (2005).

This is not the only possible response we could have. On the relativist picture, both Winifred and Lora say things that are true in their own contexts. (Since relativists normally say that truth in your own context is the relevant truth-based norm governing assertion, that is to say they don't make any assertoric mistakes.) But Winifred and Lora *disagree* because the thing Winifred says is *false* in Lora's context, and the thing that Lora says is false in Winifred's context. So from her context, Winifred can truly say that her utterance is true and Lora's false. That is to say, she can disagree with Lora, even though both of them say things that are true in their own context.

These cases provide another reason why we shouldn't adopt an indexicalist theory of conditionals, where the relevant evidence consists of the evidence that the people who hear the sentence have. If this were true, it would be a mistake for, say, Winifred to conclude that Lora's statement is false. She couldn't conclude from the fact that the antecedent of Lora's conditional might be true, but the conjunction of the antecedent and consequent is certainly false, that the conditional is false. At least, if she can conclude that, then we can't in consistency say that Lora can conclude from the fact she knows it isn't the case that the antecedent of her conditional is true and its consequent false that her statement is true. But it seems that both of these inferences are perfectly acceptable, which is impossible on this (or any other) indexicalist story.

Although I think these cases provide strong support for a relativist theory of conditionals, there is one point at which the above argument seems weak. I assumed that from the fact that Winifred and Lora are disagreeing we can conclude that not both of their utterances are true. This is too quick. Sometimes disagreement is compatible with both parties speaking truly. We'll end by looking at the possibility this is what happens in the case of conditionals.

5. Disagreement

We'll start with the discussion of disagreement in Grice (1989). Grice produced an example where people take themselves to be disagreeing even though what each of them says is *true*. In Grice's example, we imagine two people making predictions about the potential outcome of the 1966 British election. The first says, "It will be either Wilson or Heath." The second says, "No, it will be either Wilson or *Thorpe*." Grice notes that although they take themselves to be disagreeing, it is possible that what both people are saying is correct. Hence it is possible to have disagreements between two people who both utter truths. Grice takes this to support his preferred analysis of conditionals as material implications.⁵

Because of Grice's argument, we should be careful in reasoning from the existence of a disagreement to the incompatibility of the claims being asserted. But Grice's argument isn't a universal panacea. It only suggests that there *might* be something else that explains why the people take themselves to be disagreeing. If there is no explanation of the apparent disagreement other than the fact that not both of the statements are true, then we should conclude that not both of the statements are true.

In this case, the disagreement isn't too hard to find. In the example Grice discusses, it is natural to assume that the first person believes that if Wilson weren't to win, Heath would, and the second person believes that if Wilson weren't to win, Thorpe would. If we remove those implicatures, the perceived disagreement goes away. We can do this by changing the example to one set in the past tense. It is common knowledge before the election that Wilson, Heath and Thorpe are the only serious contenders. One person sees Thorpe conceding, and says "Either Wilson or Heath won." Another sees Heath conceding, and says "Either Wilson or Thorpe won." These two people do *not* take themselves to be disagreeing, and I suspect this is because there are not propositions that they either explicitly assert or tacitly endorse that are inconsistent.

_

⁵ The discussion that follows owes a lot to the discussion of Grice's examples in Stalnaker 1984: 113ff, although Stalnaker reaches rather different conclusions.

So we can only explain away Winifred and Lora's disagreement if we find propositions they either assert or implicate that are inconsistent. And in this case it doesn't seem they disagree about very much at all beyond the conditionals they utter. (Again, the important contrast here is to the case where they utter the indicative, but implicate the matching subjunctive.) So I conclude they can't both be uttering (non-relative) truths.

There are two reasons then for concluding that the truth value of indicative conditionals is relative to a context of evaluation, and in particular that it is the evidence available to the evaluator that determines the relevant epistemic context.

First, this provides the only consistent way to generalise Nolan's explanation of the plausibility of the inference patterns discussed in section two. Nolan discusses the case where the speaker uses (some of) these inferences to draw conclusions about the truth value of conditionals she utters, and his explanation of what is going on in that case is very plausible. But we can only generalise it to the case where the *hearer* uses these inferences to draw conclusions about the truth value of conditionals that other people utter if we accept a relativist theory.

Second, only the relativist can provide a plausible story about how in 'stand-off' cases of the type that Gibbard described, how the speakers can be *disagreeing* while both speaking truly. The relativist says this is possible because each of the propositions uttered is true in the context of the speaker who utters it (so they speak truly) while false in the other speakers' context (and hence a subject of disagreement). While these are not knock-down arguments for relativism, they suggest that the relativist proposal is an improvement over the existing theories concerning the truth value of indicative conditionals.