
EPISTEMIC CONDITIONALS

BY

KEN WARMBRÖD

IT has often been suggested that a speaker determines the truth or falsity of a conditional by adding the antecedent hypothetically to his stock of knowledge and checking to see whether the result entails the consequent.¹ One of the most intriguing features of this proposal is the implication that the truth-value to be assigned to a conditional depends on what a speaker knows and therefore might be different for different speakers. It is possible, of course, that the truth-value that thus gets assigned merely represents the speaker's best guess about the objective truth-value of the conditional. But a different understanding of the proposal is also possible. It could be that, for some conditionals at least, there is no objective truth-value other than the speaker-relative one assigned by the above test. I will refer to conditionals which satisfy the proposal (understood in the second way) as *epistemic conditionals* since their truth or falsity depends on what the speaker knows. As a first approximation, epistemic conditionals might be represented formally by identifying a set of sentences Γ as the speaker's body of knowledge. A particular speaker's epistemic conditional $p > q$ would then be true relative to that speaker if Γ logically implies the material conditional $p \supset q$. The concept 'speaker' here must of course be construed broadly: if I am merely contemplating asserting an epistemic conditional, presumably the truth-value is to be determined relative to my knowledge set. On the other hand, if I am evaluating a conditional statement which someone else has already asserted, the assessment should be made relative to *his* knowledge set.

The objective of this essay is to identify the class of English epistemic conditionals as precisely as possible and to characterize their truth conditions. Section I will seek to establish that there are indeed some conditionals of this type in English. Section II addresses the problem of how we can determine whether any given conditional is epistemic. Informal truth conditions are described in Section III. Finally, Sections IV and V will consider the options that are available for a more formal analysis of such conditionals.

I. Are There Epistemic Conditionals?

Certain indicative conditionals appear to conform to the epistemic model rather nicely. Consider, for example, the following variations on an example suggested by Ernest Adams.²

- (1) If Oswald did not shoot Kennedy, then someone else did.
 (2) If Oswald did not shoot Kennedy, then no one else did.

The one thing we know for sure about the Kennedy assassination is that someone shot Kennedy. Thus if Γ is our knowledge set, Γ implies the material counterpart of (1) but not that of (2). Hence, epistemic analysis gives us the right truth-values: (1) comes out true and (2) false.

But we may still be understandably reluctant to admit that any English conditionals are really epistemic. The truth-value of an epistemic conditional is speaker-relative since different speakers with different stocks of knowledge may be required to assign opposite truth-values to the same conditionals. In general, a theory of truth that recognizes objective truth-values seems preferable to one that relativizes them, unless the relativizing theory is clearly more adequate in its treatment of the linguistic phenomena.

Notwithstanding this general bias in favor of objective truth-values, I think that there are good reasons for regarding at least some English conditionals as epistemic. Suppose that a speaker S knows that Jones is a man, but S is ignorant of Jones' marital status. Then S can correctly count (3) true and (4) false.

- (3) If Jones is married, then Jones is a husband.
 (4) If Jones is married, then Jones is not a man.

The propriety of counting (4) false is supported by the observation that S should clearly consider the contrary conditional true.

If Jones is married, then Jones is a man.

Generally, if we agree that $p > q$ is true, and if p is not itself absurd, we count $p > \sim q$ false. I will refer to this condition of falsehood as the principle of *incompatible contraries* ('IC' for short).³

On the other hand, imagine that another speaker S' is certain that Jones is not a husband but does not know whether this is because Jones is unmarried or because Jones is female. Notice that what S' is assumed to know is entirely compatible with what S was supposed to know. Given his knowledge that Jones is not a husband, S' would consider (4) true. Moreover, S' would certainly want to claim that if Jones is married, then Jones is not a husband. Hence, by principle IC, S' is committed to counting (3) false. In effect, each speaker assigns truth-values exactly in accordance with the epistemic paradigm: each adds the antecedent hypothetically to what he knows and then checks to see whether the consequent follows. I do not see any reasonable prospect here for showing that either S or S' has assigned incorrect truth-values. Hence, I conclude that (3) is true and (4) is false *relative to S* and vice versa for S' .

Epistemic conditionals will often *seem* objective if enough people share the same body of background information. In the case of (1) and (2), we all share pretty much the same background information about the Kennedy assassination. Hence, we ascribe the same truth-values. But suppose that a certain individual Waldo knows that no one other than Oswald shot Kennedy, though he is not completely certain that Kennedy was in fact shot. Imagine, perhaps, that Waldo has been informed that Oswald *attempted* to kill Kennedy and that Waldo has thoroughly checked the alibi of everyone other than Oswald and verified their innocence. Knowing as he does that no one else shot Kennedy, Waldo would no doubt think that if Oswald didn't shoot Kennedy, then Kennedy wasn't shot. Hence, he would count (2) true and (1) false. It seems to me that our truth-values

are objectively no more preferable than Waldo's. The appearance of objectivity derives solely from the accident that more people share our background knowledge than share Waldo's. Things could easily have been the other way around. As in the case of *S* and *S'*, therefore, it seems natural to resolve the apparent conflict by relativizing truth-values. Relative to the knowledge set shared by most actual people, (1) is true and (2) is false. Relative to the hypothetical person Waldo, however, opposite truth-values obtain.

To dash the hope that objective truth-values may yet be saved, consider what happens when we pool our knowledge with that of Waldo. Surely, if a claim has a single "correct" truth-value, that truth-value will emerge and become more obvious as our total knowledge expands. But just the opposite happens with (1) and (2). Suppose that we know beyond any reasonable doubt that someone shot Kennedy and that no one other than Oswald shot him. Then we would know that Oswald shot Kennedy. In that case, it would be difficult to assign any truth-value at all to (1) and (2). Both conditionals are now paradoxical since part of what we know is that the antecedent is false.

How do other theories manage to avoid relativized truth-values for such conditionals? One approach is to appeal to similarity of worlds. Thus Davis maintains that an indicative conditional $p > q$ is true if the *p*-worlds (i.e. the worlds where *p* is true) that are overall most similar to the actual world are all *q*-worlds.⁴ Davis argues that the most similar worlds where Oswald did not kill Kennedy are worlds that are very much like this one both before and after the assassination. Hence, they are worlds where someone else shot Kennedy. So on Davis' analysis, (1) is objectively true and (2) is false.

This approach is attractive provided that we can agree which worlds are overall most similar to the actual world. It is enlightening to contrast Davis' intuitions about overall similarity with the views advanced by David Lewis.⁵ Lewis argues at length for a concept of similarity which implies that the worlds where Oswald did not kill Kennedy that are overall most similar to the actual world are worlds in which Kennedy was not killed. Lewis does not of course assign opposite truth-values to (1) and (2). Instead, he maintains that overall similarity of worlds is not involved in our assessment of (1) and (2).

To get objective truth-values out of an analysis based on similarity of worlds, we need to be able to agree on what is meant by 'overall similarity of worlds'. But there seems to be little prospect of showing that either Lewis or Davis is wrong about overall similarity. The problem is that the notion of world similarity itself is context dependent. The respects of similarity that are pivotal for comparisons made in one context may be quite different from the respects that are important in another context. This is obvious when things other than worlds are being compared, and there is no reason to think that worlds are a special case where respects of comparison are invariant.

Thus if we analyze conditionals such as (1) and (2) in terms of world similarity, our principal concern should be to explain how we know, for a given context, which aspects of comparison are important for that context. In Section III below, I will argue that conditionals such as (1) and (2) can indeed be analyzed in terms of world similarity. But the similarity relation that is appropriate to use depends on the speaker's body of knowledge and will thus vary from one speaker to another. Roughly speaking, the worlds most similar to the actual world are those

in which as much of what the speaker knows as possible remains true. When applied to (1) and (2), this analysis gives the same speaker-relative truth-values that I argued for above.

A different way of getting objective truth-values for such conditionals is to maintain that indicatives such as (1) and (2) are really material conditionals.⁶ The main argument for this position consists in demonstrating that the conditions for appropriately asserting an indicative conditional are the same as those for asserting the corresponding material conditional. For example, David Lewis argues that the degree of assertability of an indicative conditional $p > q$ and of its material analogue $p \supset q$ are both determined by the conditional probability $P(q/p)$. This has the surprising consequence that assertability of $p \supset q$ is *not* measured by $P(p \supset q)$ (since $P(q/p) \neq P(p \supset q)$). But Lewis offers a plausible defense of the consequence. As he puts it, "it may happen that a speaker believes a truth-functional conditional [i.e. $p \supset q$] yet he ought not to assert it."⁷ Lewis has in mind cases in which a speaker believes $p \supset q$ solely on the grounds that the antecedent is false. In such cases it would be unreasonable (misleading) to assert $p \supset q$ even though $P(\sim p)$ and hence $P(p \supset q)$ are high. So he concludes that $P(p \supset q)$ is not a measure of the assertability of the material conditional. Instead, the conditional probability $P(q/p)$ measures assertability for both the indicative and the material conditional.

But even if we agree that material and indicative conditionals have the same assertability conditions, it does not follow that they are equivalent. This becomes apparent if we shift attention from assertability conditions to a comparison of conditions of assent. It is often appropriate to assent to claims that cannot be appropriately asserted. For example, if I am reasonably certain that Oswald and no one else killed Kennedy, it is inappropriate (misleading) for me to assert the disjunction

Either Oswald shot Kennedy or no one shot Kennedy.

Such an assertion would tend to make my listeners think that I know less than I do. Nevertheless, it does not seem improper for me to *assent* to this disjunction (and hence to the corresponding material conditional) if someone else presents it and asks me whether it is true or false. However, it is not at all clear that I should assent to the corresponding *indicative* conditional (in effect (2)). My certitude about Oswald's guilt is considerably less than my confidence that Kennedy was killed. In the remote chance that Oswald did not kill Kennedy, someone else did. So I refuse assent to the indicative conditional.

There is, however, a still more telling objection to the general theory that indicatives are material. There are at least some cases in which a material conditional and its indicative counterpart agree in both their conditions for assent and assertion and yet differ in truth-values. Suppose that Smith knows that Kramer made a bet on a certain coin toss. Smith has no information about whether Kramer bet heads or tails. However, an informant whom Smith has good reason to trust has told him that the *outcome* was heads. Hence, it is appropriate for Smith to assent to and assert (5).

(5) If Kramer bet heads, then he won.

Notice, however, that though it is reasonable and appropriate for Smith to assert (5), our assumptions do not imply that his assertion would be true. The informant who told Smith that the outcome was heads has a solid record of

reliability, and so Smith has excellent inductive grounds for (5). But suppose that on this occasion the informant was lying. The outcome was really tails. In that case, I think that we would consider Smith's assertion (5) to be false. The situation is complicated at this point because there is a significant difference between what the reader knows and what Smith knows. Since the reader is privy to the fact that the outcome was tails, (5) is false relative to the reader. This is clear since the contrary of (5) must be true relative to the reader: if Kramer bet heads, he lost. Our earlier principle IC thus requires that (5) is false relative to the reader.

But let the reader now put himself back into Smith's shoes. I claim that (5) is false relative to Smith too, though not because of principle IC. Smith does not know that the outcome was tails, and so he is in no position to assert the contrary of (5). Nevertheless, he has asserted (5) because of his false belief that the outcome was heads. Surely his assertion of (5) is some sort of mistake. But what is the nature of the error? Smith's assertion cannot be considered unreasonable because it is based on good evidence. Moreover, the assertion is in no way conversationally inappropriate. (Imagine that someone has asked Smith what happened if Kramer bet heads.) The correct diagnosis of Smith's error therefore seems to be simply that Smith has asserted a falsehood.

Consider now the disjunction that corresponds to (5).

(6) Either Kramer did not bet heads or he won.

Since Smith has good evidence that the outcome was heads, he is justified in believing and asserting (6). Thus (5) and (6) are both appropriate assertions for Smith. But is (6) true? So far we have assumed nothing about what Kramer bet. Imagine, therefore, that he bet tails. Since the outcome was tails, Kramer must have won. So (6) is true. Thus we have a case of an indicative (5) and corresponding disjunction (6) both of which can be appropriately asserted and asserted to by Smith. Yet while (6) is (objectively) true, (5) seems false relative to Smith.

Such cases do not seem reconcilable to the theory that indicatives are all material. However, the explanation is straightforward if we assume that (5) is epistemic. Though the disjunction (6) is true and Smith is justified in believing it, justified true belief is not knowledge.⁸ Sentence (6) cannot be counted as part of Smith's body of knowledge because he inferred it from a falsehood, i.e. his belief that the outcome was heads. Since (6) is not something that the speaker knows, his conditional (5) is counted false under epistemic analysis. The example amounts to a conditional version of a Gettier paradox.

Such cases provide what is perhaps the strongest evidence yet that at least some English conditionals are genuinely epistemic in the sense I outlined earlier. If (5) is indeed a conditional analogue of a Gettier paradox, it shows that such conditionals are not disguised disjunctions because knowledge is not justified true belief. For the conditional (5) to be true, the corresponding disjunctive assertion must be known, not merely true, justified and conversationally appropriate.

I anticipate one significant objection to the above analysis of cases such as (5). Let us modify the example just discussed slightly. Assume that Smith still knows that Kramer has made a bet but that no one has said anything to him about the outcome. Hence he has no knowledge or belief about whether the result was tails or heads. Surely Smith should respond with 'I don't know' rather than 'false' when asked about (5). Yet my analysis requires that (5) must still be false relative to Smith's body of knowledge.⁹

The problem here is presumably not with the fact that (5) should be considered false if asserted by Smith. I take that fact to be established at this point. The problem is rather with Smith's seeming reluctance to assert that (5) is false: if it false, why should he be unwilling to claim that it is false?

I think that this modified case is easily reconciled to my analysis with the aid of one fairly obvious fact about the pragmatics of negated indicatives. In English, the assertion that an indicative conditional is false is easily confused with assertion of the contrary conditional. I do not claim, nor do I believe, that the indicative $\sim(p > q)$ is *equivalent* to $p > \sim q$. Nevertheless, there is still a reasonably high probability that a speaker who asserts the English indicative version of $\sim(p > q)$ will be interpreted by his listeners as having claimed $p > \sim q$. In the modified betting example, Smith is clearly in no position to assert either (5) or its contrary since he has no belief about the outcome. Hence, it is reasonable for him to avoid an explicit negation of (5) since such an assertion would be too readily misconstrued as a claim that the contrary of (5) is true.

The unavailability of the negation of (5) does not entail that Smith is left with no precise way of denying this conditional. Assuming that Smith has no belief whatever about the outcome and that we ask him whether Kramer won if he bet heads, Smith might well reply, 'Maybe he did, and maybe he didn't.' More precisely, Smith could be quite happy asserting a pair of indicative "might" conditionals.

(5') If Kramer bet heads, he might have won.

and

(5'') If Kramer bet heads, he might not have won.

I think that (5'') is in fact tantamount to the negation of (5). Lewis argues persuasively that a *subjunctive* "might" conditional should be analyzed as the negation of the contrary "would" conditional.¹⁰ I think that Lewis is right and that a corresponding account of indicative "might" conditionals is also correct. That is, we should interpret

If it is the case that p, then it might be the case that q.
as equivalent to

It is false that if it is the case that p, then it is not the case that q.
If this is correct, then in the modified example Smith *does* have a precise way of denying (5) without risking misconstrual. He does just that when he asserts (5''). Thus we can account for Smith's reluctance to assert the negation of (5) and we can see that he nevertheless has a non-misleading way of expressing his view that (5) is false.

II. Grammatical Features of Epistemic Conditionals

The logical and semantic concepts that logicians concoct are rarely embodied very neatly in convenient grammatical structures of English. Consider, for example, the plight of truth-functional conjunction. It is loosely tied to the English 'and', but the embodiment is an approximation at best. In English 'and' often carries a suggestion of temporal order not present in truth-functional conjunction. Many English speakers would consider the following sentences nonequivalent.

Jane got married and had a baby.

Jane had a baby and got married.

In some cases, 'and' does not even serve to conjoin sentences:

Jones stood between Smith and Kramer.

Such imperfect embodiment is also typical of conditionals. Counterfactuals, for example, have generally been characterized as conditionals whose antecedents are contrary-to-fact. But under this construal, subjunctives cannot all be counterfactuals because the antecedents of subjunctives are not always false. Moreover, future tense indicatives seem to be closer in meaning to genuine counterfactuals than to anything else. Witness the correspondence between (1') below and (1'').

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

(1') If Oswald had not shot Kennedy, someone else would have.

(1'') If Oswald does not shoot Kennedy, someone else will.

While (1) is epistemic, (1') is a clear counterfactual. Our assessment of (1') seems to be based on considerations similar to those that would have been relevant for the corresponding future indicative (1'') if it had been asserted prior to the assassination. Whatever evidence confirms or disconfirms (1') also confirms or disconfirms (1'').¹¹ Hence, while most people would consider (1) true, they would view (1') and (1'') as false.

Are there any special grammatical characteristics that can help us to identify epistemic conditionals? All the examples of epistemic conditionals in Section I were in the indicative mood and in the past or present tense. It would be convenient if all and only indicatives were epistemic. But unfortunately (1'') suggests that some indicatives (namely, future tense indicatives) are not epistemic. The suggestion is further confirmed by examples such as the following.

If this pencil is released in mid-air, it will fall.

If this pencil is accelerated to near the speed of light, its mass will approach infinity.

The past tense indicatives in the Kennedy examples (i.e., (1) and (2)) appeared at first to have objective truth-values. But I discounted this appearance on the grounds that what is general public knowledge might easily be otherwise. In the case of a future tense indicatives such as those above, however, it seems to make no difference whether the relevant body of knowledge is public information. While most people are familiar enough with gravity, the theory of relativity has still not become general public knowledge. Hence such knowledge would lead us to count the first conditional true but not the second. Nevertheless, both conditionals *are* (objectively) true. Few scientists would agree that the second conditional is false relative to those who are ignorant of relativity.

It appears, therefore, that the *combination* of the indicative mood and the past or present (but not future) tense is our best grammatical indicator that a conditional is epistemic. Subjunctive conditionals, on the other hand, are generally less amenable to epistemic interpretation (especially past tense subjunctives). It should be apparent, however, that these grammatical indicators constitute guidelines rather than exceptionless rules for determining whether an English conditional is epistemic. The lack of perfect embodiment is disconcerting, but it is at least consoling to note that we would have as much difficulty providing firm rules for deciding when an English 'and' or 'or' is to be interpreted as a truth-functional connective.

Since the grammatical indicators may not always be reliable, it is appropriate

to note that independent information about the intentions of a speaker is sometimes relevant in deciding whether a conditional is epistemic. The objective in entertaining the antecedent p of an epistemic conditional is normally to determine how one's body of knowledge would be altered if p were known. In contrast, counterfactual conditionals can be understood as assertions whose purpose is to determine how the world would be different if p were true. Thus if we know that a speaker is uncertain about p but is toying with the prospect of accepting it, we have some grounds for viewing his conditional $p > q$ as epistemic.

III. Informal Truth Conditions

I defined 'epistemic' at the beginning of the paper as follows. A conditional $p > q$ is epistemic if its truth or falsity is determined by adding p hypothetically to the speaker's stock of knowledge and determining whether the result entails q . To develop this suggestion more rigorously we will need to replace the vague notion 'stock of knowledge' with a concept that is formally more precise.

Let W be a class of worlds each of which contains the speaker and in each of which the speaker has the same psychological state (hence the same beliefs) as in the actual world. The worlds in W differ from one another in terms of which of the speaker's beliefs are true or false and in terms of other factors that affect whether those beliefs are considered knowledge. Thus a sentence p which is part of the speaker's knowledge in the actual world might be true but not counted as knowledge in some other world because the speaker's grounds in that world are in some way inadequate. For example, the belief may have been acquired in the other world via an informant who was thought to be knowledgeable but who in fact only made a lucky guess. Let K be an assignment of a knowledge set K_w of sentences of a first order language to each world w in W . For each w , K_w is the set of sentences that the speaker would know if w were actual. Roughly speaking, then, we might formulate the analysis by saying that an epistemic conditional $p > q$ is true at w (relative to knowledge assignment K) iff K_w logically implies $p \supset q$.

This formulation quickly raises a problem. As it stands, the above proposal implies that if the speaker knows that p is false, then any epistemic conditional $p > q$ will be trivially true. This may be acceptable in cases where knowledge that p is false amounts to absolute certainty that p is false (though it might be more natural to regard such a conditional as having no truth-value). But in most cases, knowing that p is false does not entail being absolutely certain that it is false, and it is this fact which raises an importunate problem for the above truth condition. What we count as knowledge is not always subjectively certain. A reasonably good history student may properly be credited with knowledge that Caligula was an emperor of Rome. Hence, he would be correct in considering an epistemic conditional such as the following to be true.

If Caligula was a vegetarian, then at least one Roman emperor was a vegetarian. But the same student's knowledge of Caligula does not prevent him from entertaining nontrivial conditionals whose antecedents deny that Caligula was an emperor. The student learned about Caligula by reading a certain history text. Hence, he correctly considers the following to be true.

If Caligula was not a Roman emperor, then my history text is wrong.

It appears that in such cases a speaker entertains an antecedent which conflicts with what he knows by tentatively dropping less certain pieces of knowledge to make his knowledge set consistent with the antecedent. More generally, a speaker determines the truth-value of $p > q$ at a world w by revising his knowledge set K_w to make it consistent with the antecedent p . The revision consists of dropping the less certain sentences until a new set K'_w consistent with p is obtained.

We can characterize this process in more formal terms by assuming that each speaker ranks the truth-functional sentences of his language in accordance with a subjective probability function P . Let us agree that P is rational at least to the extent of satisfying the laws of classical probability theory. From K_w , P produces a nested set of K_w -revisions: ϕ is a K_w -revision iff $\phi \subseteq K_w$ and every member of ϕ is more probable than any member of K_w not in ϕ . Let K'_w be the union of all K_w -revisions consistent with p . Normally, this will be the largest K_w -revision consistent with p , and hence it will be the most modest adjustment of K_w that does not trivialize the conditional. The epistemic conditional $p > q$ is then considered true at w iff K'_w implies $p \supset q$.

It must be acknowledged that there could still be problems with the above informal analysis of epistemic conditionals.¹² For example, to identify a knowledge assignment K for each speaker, we must be able to determine what knowledge should be ascribed to a person in a given psychological state under a specified set of factual circumstances. What is needed here amounts to an adequate theory of knowledge in the classical philosophical sense. It is obviously outside the scope of this essay to advance such a theory. The use of probability functions might also have to be modified. It is possible that the degree of subjective certainty of sentences in K_w is not the only factor which affects the way in which K_w is revised to accommodate p . It may well be that other factors such as informational content and explanatory power also affect which sentences are dropped and which are retained.¹³

Perhaps the best way to deal with such potential trouble spots is to formulate our analysis in slightly more general terms that allow for some differences of detail while still preserving the essential idea that the truth-value of an epistemic conditional depends on what the speaker knows. The account of truth given above could just as easily have been developed in terms of the more familiar and more general notion of world similarity. We need only to think of a speaker's knowledge and subjective priorities as determining the extent to which any world is similar to any other. The worlds closest to the actual world, for example, are those where everything the speaker knows remains true. As we drop less certain pieces of knowledge from K_w , worlds farther away come to satisfy the revised set. In general, we can say that world v is at least as similar to w as u is (i.e. $v \leq_w u$) iff every K_w -revision satisfied by u is satisfied by v . Intuitively, u will be farther away from actuality than v if one must make a more drastic modification of one's knowledge set to make it consistent with the world u than that required for world v .

I will say that v is one of the closest p -worlds to w if v is a p -world which is as close to w as any p -world. It is easily seen that v is one of the closest p -worlds to w iff v satisfies the union of K_w -revisions consistent with p (i.e. iff v satisfies K'_w). Hence, we could have formulated our truth condition above by saying that $p > q$ is true at w iff the nearest p -worlds to w are all q -worlds.¹⁴

The formulation in terms of world similarity has the virtue of generality. However, as I argued in Section I, the concept of overall similarity is at best limited in its usefulness. There are indefinitely many ways in which two worlds may be considered to resemble one another, and it is not at all clear that there is any single complex of respects of similarity that we always refer to by the phrase 'overall similarity of worlds'. To determine whether a particular speaker's conditional is true, we need empirical methods for ascertaining which respects of world similarity are important and for resolving disagreements about which of two worlds is most similar to the actual world. Fortunately, there is a reasonable prospect of resolving these difficulties in the case of epistemic conditionals if similarity relations are derived from a speaker's body of knowledge along the lines suggested earlier. The assertions that a speaker knows that p or that he is more confident about q than r are at least straightforwardly empirical claims.

One final modification of the above informal truth condition deserves consideration. It might be thought that my analysis could be simplified in a pleasing way by just abandoning person-relative truth-values in favor of subjective probabilities. After all, subjective probabilities do play a role in determining truth-values on my analysis, and such probabilities are certainly person-relative. So perhaps we could just assign a subjective conditional probability to each epistemic conditional and be done with it.¹⁵

Unfortunately this simplification would ignore the main thrust of the argument given near the end of Section I. Smith's conditional probability that Kramer won if he bet heads was very high. Nevertheless, I argued that the conditional

(5) If Kramer bet heads, then he won

should be considered false relative to Smith because Smith cannot be said to know the disjunctive analogue of (5). So whatever else we may say about these person-relative truth values, they clearly do not reduce to subjective probabilities.

IV. Epistemic Conditionals as Variably Strict

As we have seen, most epistemic conditionals are indicatives. Hence, when we begin to look for a formal analysis of epistemic conditionals, it is natural that we should turn first to proposals that have been made regarding the logic of indicatives. Let us begin with the possibility that epistemic conditionals might be variably strict. On this analysis, the truth-value of a conditional $p > q$ at a world w is relativized to a structure $\langle \leq, S \rangle$ where \leq is an assignment of a similarity relation \leq_w to each world w and S assigns a set of worlds S_w to each world w . The assignment \leq can be generated from a speaker's knowledge structure as indicated in Section III. The set S_w serves to distinguish entertainable antecedents from those that are considered absurd. The antecedent of an epistemic conditional is absurd for a given speaker if it conflicts with truths that he knows with absolute certainty (i.e. truths which he knows and which have a subjective probability of 1). Thus let ψ_w be the K_w -revision which consists of those sentences r such that $P(r) = 1$. The set S_w can then be identified with the set of worlds which satisfy every member of ψ_w . Hence, if S_w contains no p -world, p is incompatible with truths at w which are known with certainty; so p is absurd at w .

Under this type of semantics, the standard truth condition counts $p > q$ true at w if either p is absurd or the p -worlds most similar to w are all q -worlds. More precisely,

T_1 : $p > q$ is true at w relative to $\langle \leq_w, S \rangle$ iff either (i) S_w contains no p -world, or (ii) for each p -world v , if $v \geq_w u$ for every p -world u , then v is a q -world.¹⁶

The logic that results for the conditional ' $>$ ' depends on the assumptions that are made about the similarity relation. For example $A_1 - A_3$ are the assumptions of one fairly modest system which Lewis calls 'VW'.

A_1 : \leq_w is transitive

A_2 : \leq_w is strongly connected.

A_3 : w belongs to S_w and $w \leq_w v$ for every v in S_w .

It is easy to verify that where \leq and S are obtained from a knowledge assignment K and probability function P as indicated in Section III, \leq and S will satisfy $A_1 - A_3$.

The proposal that indicatives be viewed as variably strict has appeared elsewhere in slightly different versions. For example, Davis argues for such a position, but, as noted in Section I, he does not appear to think that the notion of world similarity will vary from speaker to speaker.¹⁷ Hence, Davis' analysis does not provide for the person-relative feature of epistemic conditionals.

Stalnaker has also developed a similar semantics for indicatives, and his analysis allows that the relevant respects of world similarity will vary with context.¹⁸ Stalnaker recognizes and attempts to solve a number of potential problems with this use of the semantics of variable strictness. First, some inferences that seem intuitively reasonable turn out to be invalid if the conditional is variably strict. One such illation is dubbed the 'Direct Inference' (DI):

Either the butler did it or the gardner did it.

Therefore, if the butler didn't do it, the gardner did.

A second trouble spot concerns inferences such as Hypothetical Syllogism (HS) and Contraposition (CT). These arguments are also invalid if the conditional is variably strict. However, according to Stalnaker, there are no purely indicative counterexamples to them.

Stalnaker seeks to resolve these difficulties by supplementing the logical notion of valid inference with a pragmatic theory of *reasonable inference*. The details of Stalnaker's theory need not concern us here. The important point to note is just that DI, HS and CT all turn out to be reasonable. Hence, the intuitive plausibility of these inferences is accounted for under Stalnaker's theory without making the inferences technically valid.

Unfortunately, Stalnaker's analysis depends on the claim that there are no indicative counterexamples to HS and CT, and this seems to be simply a mistake. Ernest Adams has provided numerous such counterexamples.¹⁹ All of Adams' examples are future tense indicatives and hence cannot plausibly be considered epistemic. However, it is a simple matter to construct genuine epistemic counterexamples. Returning to an illustration from Section I, S knows that Jones is a man and hence counts (3) true.

(3) If Jones is married, then Jones is a husband.

But suppose the S 's knowledge that Jones is a man is not subjectively certain: S

has no trouble entertaining the hypothesis that Jones is not a man. In that case S counts (7) true and (8) false.

(7) If Jones is married but is not really a man, then Jones is married.

(8) Therefore, if Jones is married but is not really a man, then Jones is a husband.

But (3), (7) and (8) constitute a counterexample to HS. For a counterexample to CT, we need only alter the example slightly.

If Jones is married, then Jones is not a wife.

Therefore, if Jones is a wife, Jones is not married.

It seems that Stalnaker's theory of reasonable inference would at least have to be revised to apply to epistemic conditionals since it does not distinguish acceptable cases of HS and CT from unacceptable ones. A similar problem exists for the form 'Simplification of Disjunctive Antecedents' (SDA):

$$(p \vee q) > r$$

$$\text{Therefore, } p > r.$$

Most epistemic occurrences of this inference form in ordinary speech are quite compelling. Yet, it is possible to construct counterexamples. For example, given S's confidence about Jones' masculinity, the following premise seems true relative to S.

If Jones is either a wife or a husband, Jones is a husband.

But we balk at the conclusion.

Therefore, if Jones is a wife, then Jones is a husband.

Since HS, CT and SDA are quite common in ordinary discourse, a useful theory of reasonable inference would need to distinguish systematically between the cases where these inferences are acceptable and those cases where they are not to be counted on.

Stalnaker's theory of reasonable inference does manage to explain the plausibility of the inference DI. However, when we consider the special properties of the epistemic conclusion of DI, it becomes doubtful that any supplementary theory is needed to account for this case. Note first that the discussion at the end of Section I provides a counterexample to the form of inference

$$\sim p \vee q$$

$$\text{Therefore, } p > q.$$

The disjunction (6) was objectively true. But the corresponding epistemic conditional (5) came out false when (5) was interpreted relative to Smith's body of knowledge. The inference tends to *seem* reasonable in spite of being technically invalid because the conclusion will be true relative to any speaker who *knows* that the premise is true. Thus we cannot find a single case in which we know that the premise is true and we know that the conclusion is false (relative to us).

A final problem with the view that epistemic conditionals are variably strict is that there are counterexamples to some inferences that are deemed valid under this theory. For example, consider the following instance of Substitution of Equivalent Antecedents (SEA).

$$p > q$$

$$\text{Therefore, } (p \ \& \ (r \vee \sim r)) > q.$$

Though the form is valid when '>' is variably strict the following counterexample suggests that the inference is not always reasonable.

If Jones took the 10 o'clock flight, he is now in Chicago.

Therefore, if Jones took the 10 o'clock flight and the plane either crashed or did not crash, Jones is now in Chicago.

Thus while a *prima facie* case can be made for viewing epistemic conditionals as variably strict, the analysis still has its problems. The difficulties parallel those which I have discussed elsewhere in connection with the position that counterfactual conditionals are variably strict.²⁰ There appear to be two general strategies we can follow in attempting to construct a logic for epistemic conditionals. Either we adopt a logic which is *very* weak in an effort to avoid every possible counterexample, or else we adopt a stronger logic and attempt to devise explanations for the apparent counterexamples. The liability of the former alternative is that we are left facing a vast number of apparently reasonable instances of HS, CT, SDA, and SEA that are counted technically invalid. If we wish our logic to have practical utility, we are then forced to devise some supplementary theory of reasonable inference so that the reasonable cases can be systematically distinguished from the cases where these inferences are not to be counted on. On the other hand, if we try to defend a stronger logic, we run the risk of having to explain counterexamples using *ad hoc* principles which have no purpose other than to reconcile our preferred theory to recalcitrant data. It is apparent that both approaches have their liabilities. However, I think that the prospects of success for the second approach are somewhat better. As I will argue in the next section, when a strong logical theory is outfitted with a pragmatic account of how conditionals are interpreted (which would be needed even if there were no counterexamples to be explained) the total theory provides plausible explanations for the problematic counterexamples.

V. Epistemic Conditionals as Formally Strict

When the epistemic '>' is considered to be formally strict HS, CT, SDA, and SEA are all valid. I will use Kripke's semantics and view $p > q$ as an abbreviation of $\Box(p \supset q)$. But the box ' \Box ' is not interpreted as meaning 'necessarily'. Rather, $p > q$ is true at a world w if $p \supset q$ is true at all worlds *accessible* from w . Which worlds are accessible from w will normally depend on the speaker's body of knowledge and various contextual factors. The set of worlds accessible from any given world is specified by defining an accessibility relation R . In precise terms, our new truth condition is given by T_2 .

T_2 $p > q$ is true at w under R iff, for every v such that wRv holds, $p \supset q$ is true at v under R .

Just as the semantics of variable strictness had to be supplemented with a pragmatic account of the means for identifying a similarity relation, so we now need a theory to explain how a listener knows what accessibility relation to use in interpreting any particular conditional. We begin by considering the problem of interpreting an epistemic conditional uttered more or less in isolation from any larger body of discourse. If the considerations in the section on informal truth conditions are to be adhered to, we should interpret $p > q$ in such a way that any world v will be accessible from the actual world if v is at least as close to the

actual world as the nearest worlds where p is true. Then $p > q$ will be true according to T_2 if all accessible p -worlds (i.e. the closest p -worlds are q -worlds).

To formulate this idea more precisely and generally, we again assume a pair $\langle \leq, S \rangle$ generated from a speaker's body of knowledge as described in Section IV. Let us say that an accessibility relation R results from *standard interpretation* of p relative to a structure $\langle \leq, S \rangle$ when wRv holds iff v is in S_w and $v \leq_w u$ for each p -world u .²¹ It is easily seen that, as long as we always construct a standard interpretation of an isolated conditional, this semantics gives the same truth-values as the semantics of variable strictness. What this means, of course, is just that both analyses manage to preserve the intuitive truth conditions of Section III.

The principal difference with the semantics of variable strictness arises for conditionals that are not isolated, i.e. those that occur embedded in a larger body of discourse containing (possibly) several epistemic conditionals. In the most common type of case, it seems reasonable to assume that a speaker and hearer construct a standard interpretation for the first conditional antecedent they encounter or perhaps for some hypothesis advance early in the body of discourse. Subsequent conditionals are then assessed with respect to the same accessibility relation until the discourse ends.

But this assumes that we have some way of telling when the end of a long body of discourse has been reached. Long pauses and clear topic changes are obvious signs of breaks in discourse. However, one other important way of indicating such a transition is the introduction of a new conditional which would be trivialized if interpreted relative to the old relation. If we are engaged in a conversation with S about Jones, S naturally tends to assume as much of what he knows about Jones as possible. Since S knows that Jones is a man, he asserts conditionals such as (3).

(3) If Jones is married, then Jones is a husband.

Imagine, however, that at a certain point later in the conversation someone introduces the conditional (9).

(9) If Jones is married but is not really a man, then Jones is a wife.

Though the antecedent of (9) is clearly not absurd, it does conflict with an assumption about Jones that has been operative until now in the conversation. If we continue to interpret what is said with the same assumptions in mind (hence, with the same accessibility relation), (9) will be trivialized. Given an option, we normally try to avoid trivial interpretations of what other people say. So it is clear that the old assumptions should be abandoned and a new standard interpretation formed. In effect, we avoid trivialization by treating the corpus beginning with (9) as a new body of discourse.

To express these interpretive guidelines in more precise terms, consider R to be a *normal* interpretation for a body of discourse D relative to a world w iff, for every antecedent p in D , there is some p -world v such that wRv holds. In interpreting a body of discourse D produced in the actual world, we customarily require both that the accessibility relation R result from standard interpretation early in the discourse and that R be normal for D relative to the actual world. Thus if a conditional such as (9) above is introduced which threatens to make R non-normal, we view the conditional in question as the beginning of a new body of discourse requiring a new standard interpretation.

To illustrate the use of these interpretive principles in explaining apparent counterexamples to valid inferences, consider the inference 'Construction of Disjunctive Antecedents' (CDA):

$$\begin{array}{l} p > r \\ q > r \\ \text{Therefore, } (q \vee p) > r \end{array}$$

CDA is valid regardless of whether '>' is variably strict or strict. Nevertheless, there are instances of this form which are dubious at best. The last time I checked my watch, it was about 6:30 p.m. That was at least a half hour ago. Since my watch is quite accurate, I can assert with considerable confidence (without looking at the watch of course) that

If the present time is 7 p.m., then my watch indicates 7 p.m.

But I must also admit to the following:

If my watch's battery died at exactly 7 p.m., then my watch indicates 7 p.m.

However, I am not at all happy about the conclusion required by CDA:

If either my watch's battery died at exactly 7 p.m. or the present time is 7 p.m., then my watch indicates 7 p.m.

It is not difficult to identify what it is that makes this argument suspicious. The first premise depends on a background assumption that my watch and its battery are functioning normally. I ordinarily accept the assumption and even consider it to be part of my body of knowledge. (I *know* what time it is when I look at my watch.) But this knowledge is not subjectively certain, and it is therefore easily "suspended" by the second premise. Once the assumption has been suspended, it cannot then be recovered for the conclusion. In terms of the pragmatics we have discussed, I construct a standard interpretation using the antecedent of the first premise. The assumption that the battery is functioning is true at all worlds accessible under the relation R that results from this interpretation. But this means that R will not be normal for the entire body of discourse because there are no accessible worlds where the antecedent of the second premise is true. To avoid trivialization, therefore, I automatically abandon R when the second premise is encountered and form a new interpretation and relation R' for the antecedent of the second premise. In effect, therefore, I drop the assumption that the battery is working. Since the battery is no longer functioning at all accessible worlds, the conclusion comes out false under R'. The argument *seems* invalid only because our pragmatic interpretive principles have caused us to equivocate. The most important point to note about the explanation just given is that it depends on minimal pragmatic assumptions about the way in which epistemic conditionals are interpreted, i.e. the assumption that interpretations are both standard and normal. Some such pragmatic hypotheses would be necessary in any case since the interpretation of an epistemic conditional varies with the speaker. Hence, the assumptions needed to explain the counterexample to CDA are in no sense ad hoc. Counterexamples to HS, CT, SDA, and SEA are also easily explained in terms of the concepts of standard and normal interpretation. The explanations exactly parallel those I have given elsewhere for the counterfactual versions of these inferences, so they need not be repeated here.^{22,23}

Thus I think that, for the present at least, there is a slightly better case to be made for viewing epistemic conditionals as strict than variably strict. However, I must concede that this preference arises partly from my general strategy towards the broader problems of conditional logic. Every logician must somehow strike a balance between validating as many of the intuitively reasonable inferences as possible and prohibiting inferences for which there are counterexamples. On the whole, I think that we are in a better position to deal with such problems if we adopt a relatively strong logical theory and then devise ways of accounting for the odd cases that appear unreasonable.²⁴

The University of Manitoba
Winnipeg, Manitoba, Canada

NOTES

¹The suggestion appears clearly, for example, in papers by Stalnaker and Chisholm. See Stalnaker, "A Theory of Conditionals," *American Philosophical Quarterly*, Monograph Series, no. 2 (1968), p. 101 and Chisholm, "The Contrary-to-Fact Conditional," *Mind*, 55 (1946), pp. 297-299. Both Stalnaker and Chisholm attribute the proposal to Ramsey, "General Propositions and Causality" in Ramsey's *The Foundations of Mathematics* (London: Routledge and Kegan Paul, 1931), p. 247. However, it is debatable whether Ramsey's proposal was intended to be a test for truth-values or merely for degrees of belief.

²"Subjunctive and Indicative Conditionals," *Foundations of Language*, VI (1970), pp. 89-94.

³Principle IC follows from the informal analysis of epistemic conditionals given earlier if we make the reasonable assumption that p is absurd, relative to a given speaker, in case the speaker knows for certain that p is false. A knowledge set which implies both $p \supset q$ and $p \supset \sim q$ also implies $\sim p$. So if $p > q$ is true relative to a speaker (i.e. his knowledge set implies $p \supset q$) and p is not absurd for the speaker (i.e. his knowledge set does not imply $\sim p$), then $p > \sim q$ must be false relative to the speaker (i.e. his set does not imply $p \supset \sim q$).

⁴W. Davis, "Indicative and Subjunctive Conditionals," *Philosophical Review*, LXXXVII, 4 (1979), pp. 544-564.

⁵"Counterfactual Dependence and Time's Arrow," *Noûs*, XIII, 4 (1979), pp. 455-476.

⁶The position has been defended by Lewis in "Probabilities of Conditionals and Conditional Probabilities," *Philosophical Review*, LXXXV, 3 (1976), pp. 297-315; and by Jackson in "On Assertion and Indicative Conditionals," *Philosophical Review*, LXXXVII, (1979), pp. 565-589. The position has also been criticized in Davis, *op. cit.*

⁷Lewis, *ibid.*, p. 306.

⁸This is argued persuasively in Gettier, "Is Justified True Belief Knowledge?" *Analysis*, XXIII, 6 (1963), pp. 121-123.

⁹Note that this does not conflict with principle IC because IC is not a biconditional principle. Given that p is not absurd, IC does not require that if $p > \sim q$ is false relative to a speaker, then $p > q$ will be true. It is quite possible, as in the case of Smith, for both $p > q$ and $p > \sim q$ to be false.

¹⁰Lewis, *Counterfactuals* (Cambridge, Mass.: Harvard University Press, 1973), pp. 21-24.

¹¹The existence of a correspondence between past subjunctives and future indicatives has been noted elsewhere. See, for example, Adams, *The Logic of Conditionals* (Dordrecht, Holland: D. Reidel Co., 1975), Chapter IV.

¹²Note that there is no possibility of a K_w -revision being inconsistent due to a lottery paradox. Since K_w represents what the speaker knows at world w , every member of K_w must be true at w . Hence, each K_w -revision is consistent. Note also that although K_w is presumably deductively closed, the same cannot be said for the K_w -revisions. This poses no problem for our truth condition since it is stated in terms of what the set K_w implies rather than in terms of what it contains. Finally, there could be cases in which K_w is not consistent with p . The assumption that there will not be cases of this sort is tantamount to the limit assumption. See note 14 below.

¹³Isaac Levi has suggested that such information is relevant in assessing subjunctives. It may also

be important for epistemic conditionals. See Levi, "Subjunctives, Dispositions and Chances," *Synthese*, XXXIV (1977), pp. 423–455.

¹⁴One potential source of difficulty here is the so-called 'limit assumption' which is implicit in this formulation. According to this assumption, if p is an entertainable antecedent, there must always be at least one p -world that can be deemed maximally close to the actual world. I have argued that this assumption is reasonable in "A Defense of the Limit Assumption," *Philosophical Studies*, XLII, (1982), pp. 53–66.

¹⁵Such an analysis has been proposed for indicatives by Ernest Adams in *The Logic of Conditionals*.

¹⁶This formulation is correct given the limit assumption. See note 14.

¹⁷Davis, *op. cit.*

¹⁸Stalnaker, "Indicative Conditionals" in A. Kasher, ed., *Language in Focus: Foundations, Methods and Systems* (Dordrecht, Holland: D. Reidel Co., 1976). Stalnaker develops his semantics in terms of selection functions rather than similarity relations. But this difference is incidental for our concerns and does not affect the issues raised below.

¹⁹Adams, *The Logic of Conditionals*, Chapter 1.

²⁰Warmbröd, "Counterfactuals and Substitution of Equivalent Antecedents," *Journal of Philosophical Logic*, X, 2 (1981), pp. 267–289.

²¹Here 'standard interpretation of p ' should be assumed to be defined only when p is a purely truth-functional sentence. The problems involved in interpreting nested conditionals are too complex to be addressed in the present discussion. It is easily seen then when R results from a standard interpretation and S and \leq are derived from a speaker's knowledge set as discussed in Sections III and IV, R must be reflexive. Hence our official logic for '>' should be at least as strong as the Feys-von Wright modal theory T (or M). This is consistent with the characterization of R which I argued for in "An Indexical Theory of Conditionals," *Dialogue*, XX, 4 (1981), pp. 644–664. Note in passing that principle IC discussed in Section I holds in this system because of the theorem $\sim \Box p \supset (\Box(p \supset q) \supset \sim \Box(p \supset \sim q))$.

²²See "Counterfactuals and Substitution of Equivalent Antecedents," pp. 273–287. Note that the explanation of SEA requires a "translation lore." However, the translation lore is less ad hoc when the conditional is viewed as strict because the needed translations are always done between logically equivalent formulae. Hence, the translations can be viewed as simplifying operations done prior to semantic interpretation to simplify the task of interpretation.

²³Further independent support for these pragmatic hypotheses comes from the consequences that result for epistemic conditionals whose antecedents are known for certain to be false. In such cases, there is no interpretation for the antecedent that is both standard and normal. Hence the pragmatics treats such conditionals in just the way they appear to intuition: they have no clear "correct" interpretation and hence no truth-value relative to the speaker.

²⁴I am grateful to a referee for the *Pacific Philosophical Quarterly* for helpful comments on an earlier draft.