The Philosophical Review, LXXXV, 3 (July 1976).

PROBABILITIES OF CONDITIONALS AND CONDITIONAL PROBABILITIES

DAVID LEWIS

THE truthful speaker wants not to assert falsehoods, wherefore he is willing to assert only what he takes to be very probably true. He deems it permissible to assert that A only if $P(A)$ is sufficiently close to 1, where P is the probability function that represents his system of degrees of belief at the time. Assertability goes by subjective probability.

At least, it does in most cases. But Ernest Adams has pointed out an apparent exception.[1] In the case of ordinary indicative conditionals, it seems that assertability goes instead by the conditional subjective probability of the consequent, given the antecedent. We define the conditional probability function $P(-/-)$ by a quotient of absolute probabilities, as usual:

(1) $P(C/A) =^{\text{df}} P(CA)/P(A)$, if $P(A)$ is positive.

(If the denominator $P(A)$ is zero, we let $P(C/A)$ remain undefined.) The truthful speaker evidently deems it permissible to assert the indicative conditional that if A, then C (for short, $A \rightarrow C$) only if $P(C/A)$ is sufficiently close to 1. Equivalently: only if $P(CA)$ is sufficiently much greater than $P(\bar{C}A)$.

Adams offers two sorts of evidence. There is direct evidence, obtained by contrasting cases in which we would be willing or unwilling to assert various indicative conditionals. There also is indirect evidence, obtained by considering various inferences with indicative conditional premises or conclusions. The ones that seem valid turn out to be just the ones that preserve assertability, if assertability goes by conditional probabilities for conditionals and by absolute probabilities otherwise.[2] Our judgments of validity are not so neatly explained by various rival hypotheses. In particular, they do not fit the hypothesis that the inferences that seem valid are just the ones

[1] Ernest Adams, "The Logic of Conditionals," *Inquiry*, 8 (1965), pp. 166-197; and "Probability and the Logic of Conditionals," *Aspects of Inductive Logic*, ed. by Jaakko Hintikka and Patrick Suppes (Dordrecht, 1966). I shall not here consider Adams's subsequent work, which differs at least in emphasis.

[2] More precisely, just the ones that satisfy this condition: for any positive ϵ there is a positive δ such that if any probability function gives each premise an assertability within δ of 1 then it also gives the conclusion an assertability within ϵ of 1.

that preserve truth if we take the conditionals as truth-functional.

Adams has convinced me. I shall take it as established that the assertability of an ordinary indicative conditional $A \rightarrow C$ does indeed go by the conditional subjective probability $P(C/A)$. But why? Why not rather by the absolute probability $P(A \rightarrow C)$?

The most pleasing explanation would be as follows: The assertability of $A \rightarrow C$ does go by $P(A \rightarrow C)$ after all; indicative conditionals are not exceptional. But also it goes by $P(C/A)$, as Adams says; for the meaning of $\rightarrow$ is such as to guarantee that $P(A \rightarrow C)$ and $P(C/A)$ are always equal (if the latter is defined). For short: *probabilities of conditionals are conditional probabilities.* This thesis has been proposed by various authors.[3]

If this is so, then of course the ordinary indicative conditional $A \rightarrow C$ cannot be the truth-functional conditional $A \supset C$. $P(A \supset C)$ and $P(C/A)$ are equal only in certain extreme cases. The indicative conditional must be something else: call it a *probability conditional.* We may or may not be able to give truth conditions for probability conditionals, but at least we may discover a good deal about their meaning and their logic just by using what we know about conditional probabilities.

Alas, this most pleasing explanation cannot be right. We shall see that there is no way to interpret a conditional connective so that, with sufficient generality, the probabilities of conditionals will equal the appropriate conditional probabilities. If there were, probabilities of conditionals could serve as links to establish relationships between the probabilities of nonconditionals, but the relationships thus established turn out to be incorrect. The quest for a probability conditional is futile, and we must admit that assertability does not go by absolute probability in the case of indicative conditionals.

Preliminaries

Suppose we are given an interpreted formal language equipped at least with the usual truth-functional connectives and with the

[3] Richard Jeffrey, "If" (abstract), *Journal of Philosophy*, 61 (1964), pp. 702-703; Brian Ellis, "An Epistemological Concept of Truth," *Contemporary Philosophy in Australia*, ed. by Robert Brown and C. D. Rollins (London, 1969); Robert Stalnaker, "Probability and Conditionals," *Philosophy of Science*, 37 (1970), pp. 64-80. We shall consider later whether to count Adams as another adherent of the thesis.

further connective $\rightarrow$. These connectives may be used to compound any sentences in the language. We think of the interpretation as giving the truth value of every sentence at every possible world. Two sentences are *equivalent* iff they are true at exactly the same worlds, and *incompatible* iff there is no world where both are true. One sentence *implies* another iff the second is true at every world where the first is true. A sentence is *necessary, possible,* or *impossible* iff it is true at all worlds, at some, or at none. We may think of a probability function P as an assignment of numerical values to all sentences of this language, obeying these standard laws of probability:

(2) $1 \geqslant P(A) \geqslant 0$,

(3) if A and B are equivalent, then $P(A) = P(B)$,

(4) if A and B are incompatible, then $P(A \vee B) = P(A) + P(B)$,

(5) if A is necessary, then $P(A) = 1$.

The definition (1) gives us the multiplication law for conjunctions.

Whenever $P(B)$ is positive, there is a probability function P' such that $P'(A)$ always equals $P(A/B)$; we say that P' *comes from* P *by conditionalizing on* B. A class of probability functions is *closed under conditionalizing* iff any probability function that comes by conditionalizing from one in the class is itself in the class.

Suppose that $\rightarrow$ is interpreted in such a way that, for some particular probability function P, and for any sentences A and C,

(6) $P(A \rightarrow C) = P(C/A)$, if $P(A)$ is positive;

iff so, let us call $\rightarrow$ a *probability conditional for* P. Iff $\rightarrow$ is a probability conditional for every probability function in some class of probability functions, then let us call $\rightarrow$ a *probability conditional for* the class. And iff $\rightarrow$ is a probability conditional for all probability functions, so that (6) holds for any P, A, and C, then let us call $\rightarrow$ a *universal probability conditional,* or simply a *probability conditional.*

Observe that if $\rightarrow$ is a universal probability conditional, so that (6) holds always, then (7) also holds always:

(7) $P(A \rightarrow C/B) = P(C/AB)$, if $P(AB)$ is positive.

To derive (7), apply (6) to the probability function P' that comes from P by conditionalizing on B; such a P' exists if $P(AB)$ and hence also $P(B)$ are positive. Then (7) follows by several applications of (1) and the equality between $P'(-)$ and $P(-/B)$. In the same way, if $\rightarrow$ is a probability conditional for a class of probability functions, and if that class is closed under conditionalizing, then (7) holds for any probability function P in the class, and for any A and C. (It does not follow, however, that if (6) holds for a particular probability function P, then (7) holds for the same P.)

First Triviality Result

Suppose by way of *reductio* that $\rightarrow$ is a universal probability conditional. Take any probability function P and any sentences A and C such that $P(AC)$ and $P(A\bar{C})$ both are positive. Then $P(A)$, $P(C)$, and $P(\bar{C})$ also are positive. By (6) we have:

(8) $P(A \rightarrow C) = P(C/A)$.

By (7), taking B as C or as $\bar{C}$ and simplifying the right-hand side, we have:

(9) $P(A \rightarrow C/C) = P(C/AC) = 1$,

(10) $P(A \rightarrow C/\bar{C}) = P(C/A\bar{C}) = 0$.

For any sentence D, we have the familiar expansion by cases:

(11) $P(D) = P(D/C)\cdot P(C) + P(D/\bar{C})\cdot P(\bar{C})$.

In particular, take D as $A \rightarrow C$. Then we may substitute (8), (9), and (10) into (11) to obtain:

(12) $P(C/A) = 1\cdot P(C) + 0\cdot P(\bar{C}) = P(C)$.

With the aid of the supposed probability conditional, we have reached the conclusion that if only $P(AC)$ and $P(A\bar{C})$ both are positive, then A and C are probabilistically independent under P. That is absurd. For instance, let P be the subjective probability function of someone about to throw what he takes to be a fair die, let A mean that an even number comes up, and let C mean that the six comes up. $P(AC)$ and $P(A\bar{C})$ are positive. But, *contra* (12), $P(C/A)$ is 1/3 and $P(C)$ is 1/6; A and C are not independent. More generally, let C, D, and E be possible but pairwise incompatible. There are probability functions that assign positive probability to all three; let P be any such. Let A be the disjunction $C \vee D$. Then $P(AC)$ and $P(A\bar{C})$ are positive but $P(C/A)$ and $P(C)$ are unequal.

Our supposition that $\rightarrow$ is a universal probability conditional has led to absurdity, but not quite to contradiction. If the given language were sufficiently weak in expressive power, then our conclusion might be unobjectionable. There might not exist any three possible but pairwise incompatible sentences to provide a counterexample to it. For all I have said, such a weak language might be equipped with a universal probability conditional. Indeed, consider the extreme case of a language in which there are none but necessary sentences and impossible ones. For this very trivial language, the truth-functional conditional itself is a universal probability conditional.

If an interpreted language cannot provide three possible but pairwise incompatible sentences, then we may justly call it a *trivial language*. We have proved this theorem: *any language having a universal probability conditional is a trivial language.*

Second Triviality Result

Since our language is not a trivial one, our indicative conditional must not be a universal probability conditional. But all is not yet lost for the thesis that probabilities of conditionals are conditional probabilities. A much less than universal probability conditional might be good enough. Our task, after all, concerns subjective probability: probability functions used to represent people's systems of beliefs. We need not assume, and indeed it seems rather implausible, that any probability function whatever represents a system of beliefs that it is possible for someone to have. We might set aside those probability functions that do not. If our indicative conditional were a probability conditional for a limited class of probability functions, and if that class were inclusive enough to contain any probability function that might ever represent a speaker's system of beliefs, that would suffice to explain why assertability of indicative conditionals goes by conditional subjective probability.

Once we give up on universality, it may be encouraging to find that probability conditionals for particular probability functions, at least, commonly do exist. Given a probability function P, we may be able to tailor the interpretation of $\rightarrow$ to fit.[4] Suppose that for any A and C there is some B such that $P(B/\bar{A})$ and $P(C/A)$ are equal if both defined; this should be a safe assumption when P is a probability function rich enough to represent someone's system of beliefs. If for any A and C we arbitrarily choose such a B and let $A \rightarrow C$ be interpreted as equivalent to $AC \vee \bar{A}B$, then $\rightarrow$ is a probability conditional for P. But such piecemeal tailoring does not yet provide all that we want. Even if there is a probability conditional for each probability function in a class, it does not follow that there is one probability conditional for the entire class. Different members of the class might require different interpretations of $\rightarrow$ to make the probabilities of conditionals and the conditional probabilities come out equal. But presumably our indicative conditional has a fixed interpretation, the same for speakers with different beliefs, and for one speaker before and after a change in his beliefs. Else how are disagreements about a conditional possible, or changes of mind?

[4] I am indebted to Bas van Fraassen for this observation. He has also shown that by judicious selection of the B's we can give $\rightarrow$ some further properties that might seem appropriate to a conditional connective. See Bas van Fraassen, "Probabilities of Conditionals," *Foundations of Probability and Statistics,* ed. by W. Harper and C. A. Hooker (Dordrecht, forthcoming in 1975 or 1976).

Our question, therefore, is whether the indicative conditional might have one fixed interpretation that makes it a probability conditional for the entire class of all those probability functions that represent possible systems of beliefs.

This class, we may reasonably assume, is closed under conditionalizing. Rational change of belief never can take anyone to a subjective probability function outside the class; and there are good reasons why the change of belief that results from coming to know an item of new evidence should take place by conditionalizing on what was learned.[5]

Suppose by way of *reductio* that $\rightarrow$ is a probability conditional for a class of probability functions, and that the class is closed under conditionalizing. The argument proceeds much as before. Take any probability function P in the class and any sentences A and C such that $P(AC)$ and $P(A\bar{C})$ are positive. Again we have (6) and hence (8); (7) and hence (9) and (10); (11) and hence by substitution (12): $P(C/A)$ and $P(C)$ must be equal. But if we take three pairwise incompatible sentences C, D, and E such that $P(C)$, $P(D)$ and $P(E)$ are all positive and if we take A as the disjunction $C \vee D$, then $P(AC)$ and $P(A\bar{C})$ are positive but $P(C/A)$ and $P(C)$ are unequal. So there are no such three sentences. Further, P has at most four different values. Else there would be two different values of P, x and y, strictly intermediate between 0 and 1 and such that $x + y \neq 1$. But then if $P(F) = x$ and $P(G) = y$ it follows that at least three of $P(FG)$, $P(FG)$, $P(FG)$, and $P(\overline{FG})$ are positive, which we have seen to be impossible.

If a probability function never assigns positive probability to more than two incompatible alternatives, and hence is at most four-valued, then we may call it a *trivial probability function.* We have proved this theorem: *if a class of probability functions is closed under conditionalizing, then there can be no probability conditional for that class unless the class consists entirely of trivial probability functions.* Since some probability functions that represent possible systems of belief are not trivial, our indicative conditional is not a probability conditional for the class of all such probability functions. Whatever it may mean, it cannot possibly have a meaning such as to guarantee, for all possible subjective probability functions at once, that the probabilities of conditionals equal the corresponding conditional prob-

[5] These reasons may be found in Paul Teller, "Conditionalization and Observation," *Synthese,* 26 (1973), pp. 218-258.

abilities. These is no such meaning to be had. We shall have to grant that the assertability of indicative conditionals does not go by absolute probability, and seek elsewhere for an explanation of the fact that it goes by conditional probability instead.

The Indicative Conditional as Non-Truth-Valued

Assertability goes in general by probability because probability is probability of truth and the speaker wants to be truthful. If this is not so for indicative conditionals, perhaps the reason is that they have no truth values, no truth conditions, and no probabilities of truth. Perhaps they are governed not by a semantic rule of truth but by a rule of assertability.

We might reasonably take it as the goal of semantics to specify our prevailing rules of assertability. Most of the time, to be sure, that can best be done by giving truth conditions plus the general rule that speakers should try to be truthful, or in other words that assertability goes by probability of truth. But sometimes the job might better be done another way: for instance, by giving truth conditions for antecedents and for consequents, but not for whole conditionals, plus the special rule that the assertability of an indicative conditional goes by the conditional subjective probability of the consequent given the antecedent. Why not? We are surely free to institute a new sentence form, without truth conditions, to be used for making it known that certain of one's conditional subjective probabilities are close to 1. But then it should be no surprise if we turn out to have such a device already.

Adams himself seems to favor this hypothesis about the semantics of indicative conditionals.[6] He advises us, at any rate, to set aside questions about their truth and to concentrate instead on their assertability. There is one complication: Adams does *say* that conditional probabilities are probabilities of conditionals. Nevertheless he does not mean by this that the indicative conditional is what I have here called a probability conditional; for he does not claim that the so-called "probabilities" of conditionals are probabilities of truth, and neither does he claim that they obey the standard laws of probability. They are probabilities only in name. Adams's position is therefore invulnerable to my triviality results, which were proved by

[6] "The Logic of Conditionals."

applying standard laws of probability to the probabilities of conditionals.

Would it make sense to suppose that indicative conditionals *do not* have truth values, truth conditions, or probabilities of truth, but that they *do* have probabilities that obey the standard laws? Yes, but only if we first restate those laws to get rid of all mention of truth. We must continue to permit unrestricted compounding of sentences by means of the usual connectives, so that the domain of our probability functions will be a Boolean algebra (as is standardly required); but we can no longer assume that these connectives always have their usual truth-functional interpretations, since truth-functional compounding of non-truth-valued sentences makes no sense. Instead we must choose some deductive system—any standard formalization of sentential logic will do—and characterize the usual connectives by their deductive role in this system. We must replace mention of equivalence, incompatibility, and necessity in laws (3) through (5) by mention of their syntactic substitutes in the chosen system: inter-deducibility, deductive inconsistency, and deducibility. In this way we could describe the probability functions for our language without assuming that all probabilities of sentences, or even any of them, are probabilities of truth. We could still hold that assertability goes in most cases by probability, though we could no longer restate this as a rule that speakers should try to tell the truth.

Merely to deny that probabilities of conditionals are probabilities of truth, while retaining all the standard laws of probability in suitably adapted form, would not yet make it safe to revive the thesis that probabilities of conditionals are conditional probabilities. It was not the connection between truth and probability that led to my triviality results, but only the application of standard probability theory to the probabilities of conditionals. The proofs could just as well have used versions of the laws that mentioned deducibility instead of truth. Whoever still wants to say that probabilities of conditionals are conditional probabilities had better also employ a nonstandard calculus of "probabilities". He might drop the requirement that the domain of a probability function is a Boolean algebra, in order to exclude conjunctions with conditional conjuncts from the language. Or he might instead limit (4), the law of additivity, refusing to apply it when the disjuncts A and B contain conditional conjuncts. Either maneuver would block my proofs. But if it be granted that the "probabilities" of conditionals do not obey the standard laws, I do not see what is to be gained by insisting on calling

them "probabilities". It seems to me that a position like Adams's might best be expressed by saying that indicative conditionals have neither truth values nor probabilities, and by introducing some neutral term such as "assertability" or "value" which denotes the probability of truth in the case of nonconditionals and the appropriate conditional probability in the case of indicative conditionals.

I have 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 nonexistent probabilities of truth. I have an inconclusive objection, however: the hypothesis requires too much of a fresh start. It burdens us with too much work still to be done, and wastes too much that has been done already. So far, we have nothing but a rule of assertability for conditionals with truth-valued antecedents and consequents. But what about compound sentences that have such conditionals as constituents? We think we know how the truth conditions for compound sentences of various kinds are determined by the truth conditions of constituent subsentences, but this knowledge would be useless if any of those subsentences lacked truth conditions. Either we need new semantic rules for many familiar connectives and operators when applied to indicative conditionals—perhaps rules of truth, perhaps special rules of assertability like the rule for conditionals themselves—or else we need to explain away all seeming examples of compound sentences with conditional constituents.

The Indicative Conditional as Truth-Functional

Fortunately a more conservative hypothesis is at hand. H. P. Grice has given an elegant explanation of some qualitative rules governing the assertability of indicative conditionals.[7] It turns out that a quantitative hypothesis based on Grice's ideas gives us just what we want: the rule that assertability goes by conditional subjective probability.

According to Grice, indicative conditionals *do* have truth values, truth conditions, and probabilities of truth. In fact, the indicative conditional $A \rightarrow C$ is simply the truth-functional conditional $A \supset C$. But the assertability of this truth-functional conditional does not go just by $P(A \supset C)$, its subjective probability of truth. It goes by the

[7] H. P. Grice, "Logic and Conversation," The William James Lectures, given at Harvard University in 1967.

resultant of that and something else.

It may happen that a speaker believes a truth-functional conditional to be true, yet he ought not to assert it. Its assertability might be diminished for various reasons, but let us consider one in particular. The speaker ought not to assert the conditional if he believes it to be true predominantly because he believes its antecedent to be false, so that its probability of truth consists mostly of its probability of vacuous truth. In this situation, why assert the conditional instead of denying the antecedent? It is pointless to do so. And if it is pointless, then also it is worse than pointless: it is misleading. The hearer, trusting the speaker not to assert pointlessly, will assume that he has not done so. The hearer may then wrongly infer that the speaker has additional reason to believe that the conditional is true, over and above his disbelief in the antecedent.

This consideration detracts from the assertability of $A \supset C$ to the extent that both of two conditions hold: first, that the probability $P(\bar{A})$ of vacuity is high; and second, that the probability $P(\bar{C}A)$ of falsity is a large fraction of the total probability $P(A)$ of non vacuity. The product

(13) $P(\bar{A})\cdot(P(\bar{C}A)/P(A))$

of the degrees to which the two conditions are met is therefore a suitable measure of diminution of assertability. Taking the probability $P(A \supset C)$ of truth, and subtracting the diminution of assertability as measured by (13), we obtain a suitable measure of resultant assertability:

(14) $P(A \supset C) - P(\bar{A})\cdot(P(\bar{C}A)/P(A))$.

But (14) may be simplified, using standard probability theory; and so we find that the resultant assertability, probability of truth minus the diminution given by (13), is equal to the conditional probability $P(C/A)$. That is why assertability goes by conditional probability.

Diminished assertability for such reasons is by no means special to conditionals. It appears also with uncontroversially truth-functional constructions such as negated conjunction. We are gathering mushrooms; I say to you "You won't eat that one and live." A dirty trick: I thought that one was safe and especially delicious, I wanted it myself, so I hoped to dissuade you from taking it without actually lying. I thought it highly probable that my trick would work, that you would not eat the mushroom, and therefore that I would turn out to have told the truth. But though what I said had a high subjective probability of truth, it had a low assertability and it was a misdeed to assert it. Its assertability goes not just by probability but by the

resultant of that and a correction term to take account of the pointlessness and misleadingness of denying a conjunction when one believes it false predominantly because of disbelieving one conjunct. Surely few would care to explain the low assertability of what I said by rejecting the usual truth-functional semantics for negation and conjunction, and positing instead a special probabilistic rule of assertability.

There are many considerations that might detract from assertability. Why stop at (14)? Why not add more terms to take account of the diminished assertability of insults, of irrelevancies, of long-winded pomposities, of breaches of confidence, and so forth? Perhaps part of the reason is that, unlike the diminution of assertability when the probability of a conditional is predominantly due to the improbability of the antecedent, these other diminutions depend heavily on miscellaneous features of the conversational context. In logic we are accustomed to consider sentences and inferences in abstraction from context. Therefore it is understandable if, when we philosophize, our judgements of assertability or of assertability-preserving inference are governed by a measure of assertability such as (14), that is $P(C/A)$, in which the more context-dependent dimensions of assertability are left out.

There is a more serious problem, however. What of conditionals that have a high probability predominantly because of the probability of the consequent? If we are on the right track, it seems that there should be a diminution of assertability in this case also, and one that should still show up if we abstract from context: we could argue that in such a case it is pointless, and hence also misleading, to assert the conditional rather than the consequent. This supposed diminution is left out, and I think rightly so, if we measure the assertability of a conditional $A \supset C$ (in abstraction from context) by $P(C/A)$. If A and C are probabilistically independent and each has probability .9, then the probability of the conditional (.91) is predominantly due to the probability of the consequent (.9), yet the conditional probability $P(C/A)$ is high (.9) so we count the conditional as assertable. And it does seem so, at least in some such cases: "I'll probably flunk, and it doesn't matter whether I study; I'll flunk if I do and I'll flunk if I don't."

The best I can do to account for the absence of a marked diminution in the case of the probable consequent is to concede that considerations of conversational pointlessness are not decisive. They create only tendencies toward diminished assertability, tendencies

that may or may not be conventionally reinforced. In the case of the improbable antecedent, they are strongly reinforced. In the case of the probable consequent, apparently they are not.

In conceding this, I reduce the distance between my present hypothesis that indicative conditionals are truth-functional and the rival hypothesis that they are non-truth-valued and governed by a special rule of assertability. Truth conditions plus general conversational considerations are not quite the whole story. They go much of the way toward determining the assertability of conditionals, but a separate convention is needed to finish the job. The point of ascribing truth conditions to indicative conditionals is not that we can thereby get rid entirely of special rules of assertability.

Rather, the point of ascribing truth conditions is that we thereby gain at least a prima facie theory of the truth conditions and assertability of compound sentences with conditional constituents. We need not waste whatever general knowledge we have about the way the truth conditions of compounds depend on the truth conditions of their constituents. Admittedly we might go wrong by proceeding in this way. We have found one explicable discrepancy between assertability and probability in the case of conditionals themselves, and there might be more such discrepancies in the case of various compounds of conditionals. (For instance the assertability of a negated conditional seems not to go by its probability of truth, but rather to vary inversely with the assertability of the conditional.) It is beyond the scope of this paper to survey the evidence, but I think it reasonable to hope that the discrepancies are not so many, or so difficult to explain, that they destroy the explanatory power of the hypothesis that the indicative conditional is truth-functional.

Probabilities of Stalnaker Conditionals

It is in some of the writings of Robert Stalnaker that we find the fullest elaboration of the thesis that conditional probabilities are probabilities of conditionals.[8] Stalnaker's conditional connective >

[8] "Probability and Conditionals." The Stalnaker conditional had been introduced in Robert Stalnaker, "A Theory of Conditionals," *Studies in Logical Theory,* ed. by Nicholas Rescher (Oxford, 1968). I have discussed the Stalnaker conditional in *Counterfactuals* (Oxford, 1973), pp. 77-83, arguing there that an interpretation quite similar to Stalnaker's is right for counterfactuals but wrong for indicative conditionals.

has truth conditions roughly as follows: a conditional $A > C$ is true iff the least drastic revision of the facts that would make A true would make C true as well. Stalnaker conjectures that this interpretation will make $P(A > C)$ and $P(C/A)$ equal whenever $P(A)$ is positive. He also lays down certain constraints on $P(A > C)$ for the case that $P(A)$ is zero, explaining this by means of an extended concept of conditional probability that need not concern us here.

Stalnaker supports his conjecture by exhibiting a coincidence between two sorts of validity. The sentences that are true no matter what, under Stalnaker's truth conditions, turn out to be exactly those that have positive probability no matter what, under his hypothesis about probabilities of conditionals. Certainly this is weighty evidence, but it is not decisive. Cases are known in modal logic, for instance, in which very different interpretations of a language happen to validate the very same sentences. And indeed our triviality results show that Stalnaker's conjecture cannot be right, unless we confine our attention to trivial probability functions.[9]

But it is almost right, as we shall see. Probabilities of Stalnaker conditionals do not, in general, equal the corresponding conditional probabilities.[10] But they do have some of the characteristic properties of conditional probabilities.

A possible totality of facts corresponds to a possible world; so a revision of facts corresponds to a transition from one world to another. For any given world W and (possible) antecedent A, let W_A be the world we reach by the least drastic revision of the facts of W that makes A true. There is to be no gratuitous revision: W_A may

[9] Once it is recognized that the Stalnaker conditional is not a probability conditional, the coincidence of logics has a new significance. The hypothesis that assertability of indicative conditionals goes by conditional probabilities, though still sufficiently well supported by direct evidence, is no longer unrivalled as an explanation of our judgements of validity for inferences with indicative conditional premises or conclusions. The same judgements could be explained instead by the hypothesis that the indicative conditional is the Stalnaker conditional and we judge valid those inferences that preserve truth.

[10] Although the probabilities of Stalnaker conditionals and the corresponding conditional probabilities cannot always be equal, they often are. They are equal whenever the conditional (and perhaps some non-conditional state of affairs on which it depends) is probabilistically independent of the antecedent. For example, my present subjective probabilities are such that the conditional probability of finding a penny in my pocket, given that I look for one, equals the probability of the conditional "I look for a penny > I find one." The reason is that both are equal to the absolute probability that there is a penny in my pocket now.

differ from W as much as it must to permit A to hold, but no more. Balancing off respects of similarity and difference against each other according to the importance we attach to them, W_A is to be the closest in overall similarity to W among the worlds where A is true. Then the Stalnaker conditional $A > C$ is true at the world W iff C is true at W_A, the closest A-world to W. (In case the antecedent A is impossible, so that there is no possible A-world to serve as W_A, we take $A > C$ to be vacuously true at all worlds. For simplicity I speak here only of absolute impossibility; Stalnaker works with impossibility relative to worlds.) Let us introduce this notation:

$$(15)\quad W(A) =^{\text{df}} \left\{\begin{matrix} 1 \text{ if } A \text{ is true at the world } W \\ 0 \text{ if } A \text{ is false at } W \end{matrix}\right\}.$$

Then we may give the truth conditions for non vacuous Stalnaker conditionals as follows:

$$(16)\quad W(A > C) = W_A(C), \text{ if } A \text{ is possible.}$$

It will be convenient to pretend, from this point on, that there are only finitely many possible worlds. That will trivialize the mathematics but not distort our conclusions. Then we can think of a probability function P as a distribution of probability over the worlds. Each world W has a probability $P(W)$, and these probabilities of worlds sum to 1. We return from probabilities of worlds to probabilities of sentences by summing the probabilities of the worlds where a sentence is true:

$$(17)\quad P(A) = \Sigma_W P(W) \cdot W(A).$$

I shall also assume that the worlds are distinguishable: for any two, some sentence of our language is true at one but not the other. Thus we disregard phenomena that might result if our language were sufficiently lacking in expressive power.

Given any probability function P and any possible A, there is a probability function P' such that, for any world W',

$$(18)\quad P'(W') = \Sigma_W P(W) \cdot \left\{\begin{matrix} 1 \text{ if } W_A \text{ is } W' \\ 0 \text{ otherwise.} \end{matrix}\right\}.$$

Let us say that P' *comes from* P *by imaging on* A, and call P' the *image of* P *on* A. Intuitively, the image on A of a probability function is formed by shifting the original probability of each world W over to W_A, the closest A-world to W. Probability is moved around but not created or destroyed, so the probabilities of worlds still sum to 1. Each A-world keeps whatever probability it had originally, since if W is an A-world then W_A is W itself, and it may also gain additional shares of probability that have been shifted away from $\bar{A}$-worlds. The $\bar{A}$-worlds retain none of their original probability, and gain none. All the probability

has been concentrated on the A-worlds. And this has been accomplished with no gratuitous movement of probability. Every share stays as close as it can to the world where it was originally located.

Suppose that P' comes from P by imaging on A, and consider any sentence C.

$$
\begin{aligned}
(19)\ P'(C) &= \Sigma_{W'}\, P'(W')\cdot W'(C), \text{ by (17) applied to } P';\\
&= \Sigma_{W'}\left(\Sigma_W\, P(W)\cdot\left\{\begin{matrix}1 \text{ if } W_A \text{ is } W'\\ 0 \text{ otherwise}\end{matrix}\right\}\right)\cdot W'(C), \text{ by (18);}\\
&= \Sigma_W\, P(W)\cdot\left(\Sigma_{W'}\left\{\begin{matrix}1 \text{ if } W_A \text{ is } W'\\ 0 \text{ otherwise}\end{matrix}\right\}\cdot W'(C)\right),\\
&\qquad\qquad \text{by algebra;}\\
&= \Sigma_W\, P(W)\cdot W_A(C), \text{ simplifying the inner sum;}\\
&= \Sigma_W\, P(W)\cdot W(A > C), \text{ by (16);}\\
&= P(A > C), \text{ by (17).}
\end{aligned}
$$

We have proved this theorem: *the probability of a Stalnaker conditional with a possible antecedent is the probability of the consequent after imaging on the antecedent.*

Conditionalizing is one way of revising a given probability function so as to confer certainty—probability of 1—on a given sentence. Imaging is another way to do the same thing. The two methods do not in general agree. (Example: let $P(W)$, $P(W')$, and $P(W'')$ each equal 1/3; let A hold at W and W' but not W''; and let W' be the closest A-world to W''. Then the probability function that comes from P by conditionalizing on A assigns probability 1/2 to both W and W'; whereas the probability function that comes from P by imaging on A assigns probability 1/3 to W and 2/3 to W'.) But though the methods differ, either one can plausibly be held to give minimal revisions: to revise the given probability function as much as must be done to make the given sentence certain, but no more. Imaging P on A gives a minimal revision in this sense: unlike all other revisions of P to make A certain, it involves no gratuitous movement of probability from worlds to dissimilar worlds. Conditionalizing P on A gives a minimal revision in this different sense: unlike all other revisions of P to make A certain, it does not distort the profile of probability ratios, equalities, and inequalities among sentences that imply A.[11]

Stalnaker's conjecture divides into two parts. This part is true: the probability of a nonvacuous Stalnaker conditional is the probability of the consequent, after minimal revision of the original probability

[11] Teller, "Conditionalization and Observation."

function to make the antecedent certain. But it is not true that this minimal revision works by conditionalizing. Rather it must work by imaging. Only when the two methods give the same result does the probability of a Stalnaker conditional equal the corresponding conditional probability.

Stalnaker gives the following instructions for deciding whether or not you believe a conditional.[12]

> "First, add the antecedent (hypothetically) to your stock of beliefs; second, make whatever adjustments are required to maintain consistency (without modifying the hypothetical belief in the antecedent); finally, consider whether or not the consequent is true."

That is right, for a Stalnaker conditional, if the feigned revision of beliefs works by imaging. However the passage suggests that the thing to do is to feign the sort of revision that would take place if the antecedent really were added to your stock of beliefs. That is wrong. If the antecedent really were added, you should (if possible) revise by conditionalizing. The reasons in favor of responding to new evidence by conditionalizing are equally reasons against responding by imaging instead.

Probability-Revision Conditionals

Suppose that the connective $\rightarrow$ is interpreted in such a way that for any probability function P, and for any sentences A and C,

(20) $P(A \rightarrow C) = P_A(C)$, if A is possible,

where P_A is (in some sense) the minimal revision of P that raises the probability of A to 1. Iff so, let us call $\rightarrow$ a *probability-revision conditional.* Is there such a thing? We have seen that it depends on the method of revision. Conditionalizing yields revisions that are minimal in one sense; and if P_A is obtained (when possible) by conditionalizing, then no probability-revision conditional exists (unless the language is trivial). Imaging yields revisions that are minimal in another sense; and if P_A is obtained by imaging then the Stalnaker conditional is a probability-revision conditional. Doubtless there are still other methods of revision, yielding revisions that are minimal in still other senses than we have yet considered. Are there any other methods

[12] "A Theory of Conditionals," p. 102.

which, like imaging and unlike conditionalizing, can give us a probability-revision conditional? There are not, as we shall see. The only way to have a probability-revision conditional is to interpret the conditional in Stalnaker's way and revise by imaging.

Since we have not fixed on a particular method of revising probability functions, our definition of a probability-revision conditional should be understood as tacitly relative to a method. To make this relativity explicit, let us call $\rightarrow$ a *probability-revision conditional for* a given method iff (20) holds in general when P_A is taken to be the revision obtained by that method.

Our definition of a Stalnaker conditional should likewise be understood as tacitly relative to a method of revising worlds. Stalnaker's truth conditions were deliberately left vague at the point where they mention the minimal revision of a given world to make a given antecedent true. With worlds, as with probability functions, different methods of revision will yield revisions that are minimal in different senses. We can indeed describe any method as selecting the antecedent-world closest in overall similarity to the original world; but different methods will fit this description under different resolutions of the vagueness of similarity, resolutions that stress different respects of comparison. To be explicit, let us call $\rightarrow$ a *Stalnaker conditional for* a given method of revising worlds iff (16) holds in general when W_A is taken to be the revision obtained by that method (and $A \rightarrow C$ is true at all worlds if A is impossible). I spoke loosely of "the" Stalnaker conditional, but henceforth it will be better to speak in the plural of the Stalnaker conditionals for various methods of revising worlds.

We are interested only in those methods of revision, for worlds and for probability functions, that can be regarded as giving revisions that are in some reasonable sense minimal. We have no hope of saying in any precise way just which methods those are, but at least we can list some formal requirements that such a method must satisfy. The requirements were given by Stalnaker for revision of worlds, but they carry over *mutatis mutandis* to revision of probability functions also. First, a minimal revision to reach some goal must be one that does reach it. For worlds, W_A must be a world where A is true; for probability functions, P_A must assign to A a probability of 1. Second, there must be no revision when none is needed. For worlds, if A is already true at W then W_A must be W itself; for probability functions, if $P(A)$ is already 1, then P_A must be P. Third, the method must be consistent in its comparisons. For

worlds, if B is true at W_A and A is true at W_B then W_A and W_B must be the same; else W_A would be treated as both less and more of a revision of W than is W_B. Likewise for probability functions, if $P_A(B)$ and $P_B(A)$ both are 1, then P_A and P_B must be the same.

Let us call any method of revision of worlds or of probability functions *eligible* iff it satisfies these three requirements. We note that the methods of revising probability functions that we have considered are indeed eligible. Conditionalizing is an eligible method; or, more precisely, conditionalizing can be extended to an eligible method applicable to any probability function P and any possible A. (Choose some fixed arbitrary well-ordering of all probability functions. In case P_A cannot be obtained by conditionalizing because $P(A)$ is zero, let it be the first, according to the arbitrary ordering, of the probability functions that assign to A a probability of 1.) Imaging is also an eligible method. More precisely, imaging on the basis of any eligible method of revising worlds is an eligible method of revising probability functions.

Our theorem of the previous section may be restated as follows. *If $\rightarrow$ is a Stalnaker conditional for any eligible method of revising worlds, then $\rightarrow$ is also a probability-revision conditional for an eligible method of revising probability functions; namely, for the method that works by imaging on the basis of the given method of revising worlds.* Now we shall prove the converse: *if $\rightarrow$ is a probability-revision conditional for an eligible method of revising probability functions, then $\rightarrow$ is also a Stalnaker conditional for an eligible method of revising worlds.* In short, the *probability-revision conditionals are exactly the Stalnaker conditionals.*

Suppose that we have some eligible method of revising probability functions; and suppose that $\rightarrow$ is a probability-revision conditional for this method.

We shall need to find a method of revising worlds; therefore let us consider the revision of certain special probability functions that stand in one-to-one correspondence with the worlds. For each world W, there is a probability function P that gives all the probability to W and none to any other world. Accordingly, by (17),

$$(21)\quad P(A) = \left\{ \begin{array}{l} 1 \text{ if } A \text{ is true at } W \\ 0 \text{ if } A \text{ is false at } W \end{array} \right\} = W(A)$$

for any sentence A. Call such a probability function *opinionated,* since it would represent the beliefs of someone who was absolutely certain that the world W was actual and who therefore held a firm opinion about every question; and call the world W where P concentrates all

the probability the *belief world of P*.

Our given method of revising probability functions preserves opinionation. Suppose P were opinionated and P_A were not, for some possible A. That is to say that P_A gives positive probability to two or more worlds. We have assumed that our language has the means to distinguish the worlds, so there is some sentence C such that $P_A(C)$ is neither 0 nor 1. But since P is opinionated, $P(A \rightarrow C)$ is either 0 or 1, contradicting the hypothesis that $\rightarrow$ is a probability-revision conditional so that $P_A(C)$ and $P(A \rightarrow C)$ are equal.

Then we have the following method of revising worlds. Given a world W and possible sentence A, let P be the opinionated probability function with belief world W, revise P according to our given method of revising probability functions, and let W_A be the belief world of the resulting opinionated probability function P_A. Since the given method of revising probability functions is eligible, so is this derived method of revising worlds.

Consider any world W and sentences A and C. Let P be the opinionated probability function with belief world W, and let W_A be as above. Then if A is possible,

$$\begin{aligned} (22)\quad W(A \rightarrow C) &= P(A \rightarrow C), \text{ by (21);} \\ &= P_A(C), \text{ by (20);} \\ &= W_A(C), \text{ by (21) applied to } W_A. \end{aligned}$$

So $\rightarrow$ is a Stalnaker conditional for the derived method of revising worlds. *Quod erat demonstrandum.*[13]

Princeton University

[13] An earlier version of this paper was presented at a Canadian Philosophical Association colloquium on probability semantics for conditional logic at Montreal in June 1972. I am grateful to many friends and colleagues, and especially to Ernest Adams and Robert Stalnaker, for valuable comments.