

## PROBABILITIES OF CONDITIONALS AND CONDITIONAL PROBABILITIES II

David Lewis

Adams's thesis about indicative conditionals is that their assertability goes by the conditional subjective probability of the consequent given the antecedent, in very much the same way that assertability normally goes by the subjective probability of truth.<sup>1</sup> The thesis is well established; the remaining question is how it may best be explained. The nicest explanation would be that the truth conditions of indicative conditionals are such as to guarantee the equality

$$(*) \quad P(A \rightarrow C) = P(C/A) =^{\text{df}} P(CA)/P(A)$$

whenever  $P(A)$  is positive. In a previous paper,<sup>2</sup> I argued that this nicest explanation cannot be right. After reviewing my previous argument, I shall here extend it in order to plug some loopholes.

### I

I began with a *first triviality result*, as follows. Except in a trivial case, there is no way to interpret  $\rightarrow$  uniformly<sup>3</sup> so that  $(*)$  holds

---

<sup>1</sup>See Ernest W. Adams, "The Logic of Conditionals," *Inquiry* 8 (1965), pp. 166–197; *The Logic of Conditionals* (Dordrecht: Reidel, 1975).

<sup>2</sup>"Probabilities of Conditionals and Conditional Probabilities," *The Philosophical Review* 85 (1976), pp. 297–315; reprinted (with postscripts) in my *Philosophical Papers*, Volume II (Oxford: Oxford University Press, 1986).

<sup>3</sup>I take it to be part of the "nicest explanation" now under attack that  $\rightarrow$  is to be interpreted uniformly, the same way in the context of one probability function as in the context of another. The proposal that  $\rightarrow$  has a non-uniform interpretation, so that for each  $P$  we have a  $\rightarrow_P$  that satisfies  $(*)$ , but in general a different  $\rightarrow_P$  for different  $P$ , is a rival hypothesis, an alternative way to explain Adams's thesis. It is unscathed by my arguments here and in the previous paper. See Robert Stalnaker, "Indicative Conditionals," *Philosophia* 5 (1975), pp. 269–286, reprinted in W. L. Harper *et al.*, eds., *Ifs* (Dordrecht: Reidel, 1981); and Bas van Fraassen, "Probabilities of Conditionals," in W. L. Harper and C. A. Hooker, eds., *Foundations of Probability Theory, Statistical Inference and Statistical Theories of Science*, Volume I, (Dordrecht: Reidel, 1976).

universally, for every probability function  $P$  (as well as for every antecedent  $A$  and consequent  $C$ ).

But you might say that (\*) doesn't have to hold universally, throughout the class of all probability functions whatever. To explain Adams's thesis, it is good enough if (\*) holds only throughout the class of *belief functions*: probability functions that represent possible systems of belief.

I agree. However I cited 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; I concluded that the class of belief functions is closed under conditionalizing; and I appealed to a *second triviality result*, as follows. Except in a trivial case, there is no way to interpret  $\rightarrow$  uniformly so that (\*) holds throughout a class of probability functions closed under conditionalizing.

In the previous paper, that was the end of my argument. But I now think the end came too soon. I owe you answers to two more objections.

## II

You might say that not just any proposition<sup>4</sup> could be an item of evidence. Even granted that change of belief takes place (at least sometimes) by conditionalizing on one's total evidence, that does not mean that the class of belief functions must be closed under conditionalizing generally. It is enough that the class should be closed under conditionalizing on those propositions that could be someone's total evidence.

In proving the second triviality result, I did not use the full strength of my assumption that the class in question was closed under conditionalizing. I only conditionalized twice: once on a proposition  $C$ , once on its negation  $\neg C$ . But that, you might well say, was bad enough. A proposition that could be someone's total

---

<sup>4</sup>Here I switch from sentences (in the previous paper) to propositions as the bearers of subjective probability. The reason is that our items of evidence might be propositions that have no accurate expression in any language we are capable of using.

evidence must be, in certain respects, highly specific. But to the extent that a proposition is specific, its negation is unspecific. So one time or the other, whether on  $C$  or on  $\neg C$ , I conditionalized on a proposition of the wrong sort; and the result may well have been a probability function that was not a belief function.

I agree. I did not give any good reason why belief functions should be closed under conditionalizing generally. I should only have assumed that they are closed under conditionalizing on a certain special class of *evidence propositions*.

Consider a limited class of evidence propositions: those that characterize the total evidence available to a particular subject at a particular time, as it is or as it might have been, in a maximally specific way. These propositions—as always with the maximally specific alternatives for a single subject matter—comprise a partition: they are mutually exclusive and jointly exhaustive. Further, the subject's limited powers of discrimination will ensure that this partition is a finite one.

So the hypothesis now before us says that (\*) holds throughout the class of belief functions, and that this class is closed under conditionalizing on the members of a certain finite partition of evidence propositions. Against this I appeal to a *third triviality result*, as follows. Except in a trivial case, there is no way to interpret  $\rightarrow$  uniformly so that (\*) holds throughout a class of probability functions closed under conditionalizing on the propositions in some finite partition.

The proof differs little from my previous proofs of the first and second results. Let  $P$  be a probability function in the class. Let  $C, D, \dots$  be all the propositions in the partition to which  $P$  assigns positive probability; we assume that there are at least two such propositions. Let  $A$  be a proposition such that  $P(A/C), P(A/D), \dots$ , are all positive, and such that  $P(C/A) \neq P(C)$ . If there are no such  $P, C, D, \dots$ , and  $A$ , that is the case I am here calling 'trivial'. We may be sure that the case of the class of belief functions and a partition of evidence propositions will not be thus trivial.

By finite additivity, the definition of conditional probability, and the incompatibility of  $C, D, \dots$ , we have that

$$P(A \rightarrow C) = P(A \rightarrow C/C)P(C) + P(A \rightarrow C/D)P(D) + \dots$$

Suppose for *reductio* that (\*) holds throughout the class. By applying it in turn to  $P$  and to all the functions  $P(-/C)$ ,  $P(-/D)$ , . . . that come from  $P$  by conditionalizing on  $D$ ,  $C$ , . . . respectively, we have

$$\frac{P(CA)}{P(A)} = \frac{P(CA/C)P(C)}{P(A/C)} + \frac{P(CA/D)P(D)}{P(A/D)} + \dots$$

The first term of the right-hand sum simplifies and the other terms vanish, so we have

$$P(C/A) = P(C),$$

which contradicts our choice of  $A$  and thereby completes the *reductio*.

### III

So if you grant that change of belief takes place (at least sometimes) by conditionalizing, and only insist that the conditionalizing must be on evidence propositions, then I take your point but I can make my case against (\*) despite it. However you might go further. You might say that conditionalizing never yields one belief function from another.

Why? You might say it because you run my own argument in reverse.<sup>5</sup> Adams's thesis is compelling; it needs to be explained; it turns out that even a modicum of closure under conditionalizing would wreck the explanation; therefore even that modicum of closure is to be rejected. I reply that if (\*) does not hold throughout the belief functions, that wrecks only one explanation of Adams's thesis. Other hypotheses remain.<sup>6</sup>

You might better say it for a more direct reason.<sup>7</sup> If one belief

---

<sup>5</sup>Thus Wayne Davis, in discussion.

<sup>6</sup>Of several available proposals to explain Adams's thesis without the aid of (\*), I now favor that of Frank Jackson, "On Assertion and Indicative Conditionals," *The Philosophical Review* 87 (1979), pp. 565–589. In particular, I take it to be better (for the reasons Jackson gives) than the explanation of Adams's thesis that I proposed in "Probabilities of Conditionals and Conditional Probabilities."

<sup>7</sup>Thus Anthony Appiah, "The Importance of Triviality," *The Philosophical Review* 95 (1986), pp. 209–231. But Appiah is no upholder of (\*),

function came from another by conditionalizing on the evidence, it would be *irregular*. It would not spread its probability over all the possibilities there are, but would assign probability zero to all possibilities that conflict with the evidence. You might insist that an irregular probability function cannot represent a reasonable system of belief. It would be overconfident to respond to evidence by conditionalizing and thereby falling into irregularity. To do so is to dismiss the genuine possibility that one has mistaken the evidence. It is to be ready to bet against that possibility at any odds, no matter how high the stakes. And it is to dismiss the possibility forever, no matter how much later evidence may call for reconsideration. For once a proposition gets probability zero, no subsequent conditionalizing ever can restore it to positive probability. (And neither can the generalized conditionalizing we shall soon consider.) Surely it would be more reasonable to respond to evidence in a slightly more moderate way, almost conditionalizing but not quite. If a possibility conflicts with the evidence, by all means bring its probability down very low—but never quite to zero.

It's one thing to say, as you have, that an irregular probability function cannot represent a reasonable system of belief; it's something else to say that it cannot represent a system of belief at all. The latter is what you need if, despite my triviality results so far, you still say that (\*) holds throughout the class of all belief functions. But you can cross the gap by appealing to a theory of intentionality according to which—roughly speaking—the content of a total mental state is the system of belief and desire that best rationalizes the behavior to which that state would tend to dispose one. How could an unreasonable belief function ever be part of the system that best rationalizes anything? If it cannot, it never is suited to be part of the content of any mental state, in which case it is not a belief function after all. This theory of reason as constitutive of content must be handled with care, lest it prove too much—namely, that there is no such thing as unreason—but it may well be that a sufficiently qualified form of it would meet your needs.

---

although he faults my argument against it. He rejects it on the strength of an argument advanced by I. F. Carlstrom and C. S. Hill in their review of Adams, "The Logic of Conditionals," *Philosophy of Science* 45 (1978), pp. 155–158.

(What if some state would dispose one to behavior that would indeed suggest a complete and irrevocable dismissal of the possibility that one has mistaken the evidence? Even so, it is not clear that an irregular belief function is called for. If a system is to do well at rationalizing behavior, of course one *desideratum* is that the behavior should fit the prescriptions of the system. But another *desideratum*, you may plausibly say, is that the system itself should be reasonable. Even if an irregular function does best on fit, a not-quite-irregular function may yet do better on balance.)

For what it is worth, I would insist that the ideally rational agent does conditionalize on his total evidence, and thereby falls into irregularity. He never does mistake the evidence, wherefore he may and he must dismiss the possibility that he has mistaken it. Else there is a surefire way to drain his pockets: sell him insurance against the mistakes he never makes, collect the premium, never pay any claims.<sup>8</sup> This surprising conclusion does nothing to show that we, the imperfect folk who do run the risk of mistaking our evidence, would be reasonable to conditionalize ourselves into irregularity. Not at all—rather, it shows that we differ more from the ideally rational agent than might have been thought. After all, many of our virtues consist of being good at coping with our limitations. (Whatever the divine virtues may be, they can scarcely include curiosity, fortitude, or gumption.) Likewise our cautious refusal to conditionalize quite all the way helps us to cope with our fallibility. One departure from the perfection of ideal rationality demands another.

(If I am right that the ideally rational agent conditionalizes, then I have shown at least that (\*) does not hold throughout the class of *his* belief functions. So if Adams's thesis applies to his indicative conditionals, it must be otherwise explained. But what does that

---

<sup>8</sup>In other words, if the ideally rational agent did not conditionalize, and reserved some probability for the hypothesis that he had mistaken his evidence, then he would be vulnerable to a "diachronic Dutch book"—which, by definition of ideal rationality, he is not. See Paul Teller, "Conditionalization and Observation," *Synthese* 26 (1973), pp. 218–258, and Bas van Fraassen, "Belief and the Will," *Journal of Philosophy* 81 (1984), pp. 235–256, for discussion of diachronic Dutch books. However, Teller and van Fraassen would accept the conclusion I draw only in much-qualified form, if at all.

show about us? You might well doubt whether Adams's thesis does apply to the ideally rational agent, given that we differ from him more than might have been thought.)

But if we are talking of everyday, less-than-ideal rationality, then I agree that it is unreasonable to conditionalize, reasonable to shun irregularity. Nor can I challenge your further conclusion that no belief function—anyhow, none suited to ordinary folk—is irregular. I accept it, if not wholeheartedly, at least for the sake of the argument.

If we do not conditionalize on our evidence, what do we do instead? To this question, Jeffrey has offered a highly plausible answer.<sup>9</sup> In conditionalizing, the probability of the evidence proposition is exogenously raised to one; the probability of anything else conditional on that evidence is left unchanged. Jeffrey generalizes this in a natural way, as follows. The probability of the evidence proposition is exogenously raised, not to one but to some less extreme value; the probability of the negation of the evidence proposition is lowered by an equal amount, as must be done to preserve additivity; and the probabilities of anything else conditional on the evidence and conditional on its negation both are left unchanged. So if  $P$  is the original probability function and the probability of proposition  $C$  is exogenously raised by amount  $x$ , we get a new probability function  $P_x$  given by the schema

$$(GC) \quad P(B) = P_x(B) + x[P(B/C) - P(B/-C)]$$

whenever  $P(C)$  and  $P(-C)$  and  $x$  are positive,  $C$  is an evidence proposition, and  $x$  is less than  $P(-C)$ . (The final restriction excludes the case that  $x = P(-C)$ , which reduces to ordinary conditionalizing on  $C$  and engenders irregularity; and also the case that  $x$  is greater than  $P(-C)$ , in which  $P_x$  would assign negative values and so would not be a genuine probability function.) I shall assume that

---

<sup>9</sup>Richard C. Jeffrey, *The Logic of Decision* (New York: McGraw Hill, 1965, and Chicago: University of Chicago Press, 1983), Chapter 11. Jeffrey actually describes a more general case, in which there may be exogenous change to the probabilities of several evidence propositions simultaneously, but for the argument to follow it is enough to consider the simpler case. For further discussion, see Paul Teller, *op. cit.*

change of belief under the impact of evidence takes place (at least sometimes) in conformity to schema (GC).

So the hypothesis now before us says that (\*) holds throughout the class of belief functions, and that this class is closed under change in conformity to schema (GC). Against this I appeal to a *fourth triviality result*, as follows. Except in a trivial case, there is no way to interpret  $\rightarrow$  uniformly so that (\*) holds throughout a class of probability functions closed under change in conformity to (GC).

Let  $P$  be a probability function in the class. Let  $C$  be an evidence proposition such that  $P(C)$  and  $P(-C)$  are positive. Let  $A$  be a proposition such that  $P(A/C)$  and  $P(A/-C)$  are positive and unequal. If there are no such  $P$ ,  $C$ , and  $A$ , that is the case I am now calling 'trivial'. We may be sure that the class of belief functions will not be trivial in this way.

Suppose for *reductio* that (\*) holds throughout the class, hence both for  $P$  and for any  $P_x$  with  $x$  between 0 and  $P(-C)$ . So we have

$$P(A \rightarrow C) = P(CA)/P(A),$$

$$P_x(A \rightarrow C) = P_x(CA)/P_x(A);$$

where by (GC),

$$P_x(A \rightarrow C) = P(A \rightarrow C) + x[P(A \rightarrow C/C) - P(A \rightarrow C/-C)],$$

$$P_x(CA) = P(CA) + x[P(CA/C) - P(CA/-C)]$$

$$= P(CA) + xP(A/C),$$

$$P_x(A) = P(A) + x[P(A/C) - P(A/-C)].$$

From these equations, we can derive that whenever  $x$  is between 0 and  $P(-C)$ ,

$$\begin{aligned} & x[P(A/C) - P(A/-C)][P(A \rightarrow C/C) - P(A \rightarrow C/-C)] \\ &= P(A/C) - P(A)[P(A \rightarrow C/C) - P(A \rightarrow C/-C)] - \\ & \quad P(C/A)[P(A/C) - P(A/-C)]. \end{aligned}$$

The only way this can hold for a range of values of  $x$  is that the coefficient of  $x$  on the left hand side is zero, and the right hand side also is zero. There are two cases.



Case 1. The coefficient is zero because  $P(A/C) = P(A/-C)$ . But that contradicts our choice of  $A$ .

Case 2. The coefficient is zero because  $P(A \rightarrow C/C) = P(A \rightarrow C/-C)$ . Then since the right hand side also is zero, and since its second term vanishes, we have that

$$\begin{aligned} 0 &= P(A/C) - P(C/A)[P(A/C) - P(A/-C)] \\ &= P(A/C)[1 - P(C/A)] + P(C/A) P(A/-C) \\ &= P(A/C) P(-C/A) + P(C/A) P(A/-C), \end{aligned}$$

which contradicts our choice of  $P$ ,  $C$ , and  $A$ , whereby all of  $P(A/C)$ ,  $P(-C/A)$ ,  $P(C/A)$ , and  $P(A/-C)$  must be positive. This completes the *reductio*.

*Princeton University*