Adams's Puzzle about Counterfactuals
Dorothy Edgington

1. *Prologue: What we owe to Ernie Adams*

It is a particular pleasure to be speaking at this memorial session for Ernest Adams, to whom I owe an enormous amount philosophically. Some of the virtues of his approach to philosophy will emerge in the body of the talk. But there are a couple of things much to his credit that I thought I'd mention at the outset, although they don't figure in this talk.

First, there's Adams's probabilistic theory of deductive validity, which is beautiful, simple, and of considerable interest and value independently of its application to conditionals. Take an argument that consists only of propositions. He showed that all and only necessarily truth-preserving inferences are probability-preserving in the following sense: the uncertainty (1 minus the probability) of the conclusion cannot exceed the sum of the uncertainties of the premises. i.e. there is no probability distribution in which the uncertainty of the conclusion exceeds the sum of the uncertainties of the premises. An equivalent result: all and only necessarily truth-preserving arguments have this property: pick any value short of 1 for the conclusion, then there is a value short of 1 for the premises such that if all the premises have that, they guarantee at least the value for the conclusion. We need these results because we need an answer to the question: what's the good of knowing an argument is valid when I'm not completely certain that its premises are true? Without that result, deductive logic is of very limited value when applied to ordinary contingent statements. The result also goes a long way towards defusing the lottery paradox and the like.

Adams extended the probabilistic criterion to arguments with conditionals. His attitude in his early work was something like this: no one has been able to figure out satisfactory truth conditions for conditionals, but we have a pretty good tool for assessing them probabilistically—the notion of conditional probability. So the valid arguments should be those which preserve probability and conditional probability in the way he had developed. And a nice logic emerged.

Second: Adams was responsible for shaping a logic of conditionals which is widely accepted now, even by many who do not accept his interpretation of conditionals. He published two important papers in 1965 and 1966 ('A Logic of Conditionals', *Inquiry* 1965; 'Probability and the Logic of Conditionals' in Hintikka and Suppes (eds) *Aspects of Inductive Logic* 1986), in which, inter alia, he gave the counterexamples to transitivity, strengthening and contraposition which later became famous in the work of Stalnaker and Lewis. (Stalnaker and Lewis, who were about 24 at the time, had yet to begin their work on conditionals.) For instance he was the first to give examples like this:

If Brown wins the election, Smith will resign immediately afterwards.
If Smith dies before the election, Brown will win.
Therefore, if Smith dies before the election, Smith will resign immediately afterwards.

(Adams 1965, p. 166).

(The premises may get assigned e.g. 99%, 1 and 0, so they fail his criterion.)

Stalnaker, who was the first to give a possible-worlds semantics for conditionals, was directly aiming to give truth conditions which accord with Adams's logic and the idea (found originally in Ramsey) that our confidence in a conditional is the conditional probability of consequent given antecedent. That attempt had an unhappy ending. That did not surprise Adams: he had expressed scepticism about truth conditions for conditionals so evaluated in these early papers of 1965 and 1966. Indeed he did have a Lewis-style proof of his own, but did not publish it. (Nor, incidentally, would it have surprised Ramsey: there are indications in 'General Propositions and Causality', the paper with the famous footnote, that he realised that this was of thinking of conditionals is incompatible with thinking of them as propositions. The last sentence of that paper begins 'The difficulty comes fundamentally from taking every sentence to be a proposition'.)

Adams's influence on Lewis is harder to trace than his influence on Stalnaker, but Lewis undoubtedly had read those early papers, and it is possible that they provided some inspiration for his logic of counterfactuals. And although Lewis did not go all the way with Adams on indicatives, he and Jackson go part of the way: in his 1976 Lewis writes 'Adams has convinced me. I shall take it as established that the assertability of the ordinary indicative conditional A-->C does indeed go be the conditional subjective probability p(C | A)'. (p. 134)

2. *Some background and **some** motivation for thinking of conditional judgements in terms of conditional probabilities*. Adams takes uncertainty seriously. The non-trivial conditional judgements that matter to us are very often uncertain. Will he recover if he has the operation? Will the dog attack me if I approach? If I pick a red ball, will it have a black spot? The corresponding judgements, 'He will recover if he has the operation', 'The dog will attack if you approach', etc. are rarely judgements of which one can be certain. We can judge them to be more or less probable. The same applies when time passes, the antecedents prove false, and we reiterate in the (so-called) subjunctive mood: 'He would have recovered if he had had the operation'; 'The dog would have attacked me if I had approached'; and so on.

It is undeniable that such judgements are often uncertain. When it is taken for granted that the name of the game is to find truth conditions for conditionals, the circumstances in which a conditional judgement is true, it is easy to ignore uncertain conditional judgements, and to test one's theory with examples in which someone would claim to know that if A, C. You don't have to deny that there are uncertain conditional judgements, but they are not your special business—they are the business of a general theory of uncertainty about what is true. They are not in focus. But when we do focus on uncertain conditional judgements, we find that truth conditional theories have bad consequences. For instance, the truth-functional theory has the consequence that all conditionals whose antecedents are unlikely, are likely too be true. And the most popular truth conditions for counterfactuals—that the consequent C is true in *all* relevant A-worlds—brings far too many out to be certainly knowably false.

When we do focus on uncertain conditional judgements, conditional probability leaps to mind as a very natural way to evaluate them. 'I'm around 90% certain that you'll be cured if you have the operation'; i.e., 'I'm about 90% certain that you'll be cured under the supposition that you have the operation'; i.e., 'It's about 9 times more likely

that you'll have the operation and be cured, than it is that you'll have the operation and not be cured'; all seem equivalent.

If we want to preserve the easy transition to the counterfactual, we also want (in cases where there is no change in evidence) that it is 90% likely that he would have been cured if he had had the operation; and so on. (Those who follow Adams on indicatives and Lewis on counterfactuals agree with the first 90%, but then say it is certainly false that he would have been cured if he had had the operation. And that is an unfortunate combination.) So we want the counterfactuals to be assessed by a conditional probability too, *but* it is typically a conditional probability distanced (e.g. temporally) from your present belief state: you now think that, at the earlier time when there was still some chance that he would have the operation, it was 90% likely that he would have been cured, if he had done so. That, roughly, was Adams's idea.

3. *Adams on counterfactuals and their connection with Bayesian reasoning*. The fourth and final chapter of Adams's *The Logic of Conditionals* is about counterfactuals. While his work on indicatives is well known and accepted by some prominent philosophers such as Gibbard and Bennett, his work on counterfactuals has been almost entirely ignored. A possible explanation of this (though not one to which I give much credence) is that section 8 of this chapter, headed 'A counterexample', presents (what he takes to be) a counterexample to his own theory, one for which he does not find a satisfactory solution. (I have met the reaction once, from a student 'We don't need to bother with Adams's theory of counterfactuals because he himself has refuted it'.) Skyrms is the only philosopher I know of who takes up the matter, and I'll comment on Skyrms later. My aim is to defuse the counterexample, i.e. to solve the puzzle that generates it in a way consistent with Adams's theory.

Adams's hypothesis about counterfactuals is that they involve what he calls an 'epistemic past tense'. It's not *altogether* clear what that phrase means, but other instances would be 'It was to be expected that' or 'They should have been here by now' (not a moral should—an epistemic should). The idea is that the (conditional) probability to be attached at the time of utterance to 'If he had had the operation, he would have been cured' is that which you now endorse for the earlier indicative 'If he has the operation he will be cured'.

It may be insufficiently general to focus on time, because not all conditionals involve specific times, but there may be other ways, other than temporal ones, of distancing yourself from your present epistemic state, and I'll leave this aside here. I want to mention another feature of Adams's discussion.

He considers it important to address the question: what are counterfactuals for? What do they do for us? This should throw some light on the question of why we evaluate them the way we do. While this does not pretend to be an exhaustive answer, he points out that they play an essential role in non-demonstrative inferences to conclusions about what is actually the case. He mentions two kinds of such inference, which he calls 'counterfactual modus tollens' and 'inference to the best explanation', but on closer inspection they can be seen as different sides of the same coin.

(1) You are driving, of an evening, in the dark, close to the house of some friends, and have considered paying a visit. You turn the corner. 'They're not at home', you say, 'for the lights are off. And if they had been at home the lights would have been on'.
(2) A patient is brought to the hospital in a coma. 'I think he must have taken arsenic', says the doctor, after examination, 'for he has [such-and-such] symptoms. And these are just the symptoms he would have if he had taken arsenic'.

Note that in a way it is rather odd to call the conditional is the second case a 'counterfactual', for the doctor is arguing for the *truth* of the antecedent! But there is a certain sense in which, in using the 'would'-conditional (as he must) the doctor is abstracting (distancing himself) from his actual belief state, in that he is ignoring the fact that he knows that the patient *does* have these symptoms. So, in a sense, he is appealing to a counter-actual belief state.

These should be seen as (at best) defeasible inferences. The second is no good if there are some conditions other than having taken arsenic which would also produce these symptoms. That is, it is a good inference only to the extent that the doctor also has grounds for thinking that if he had not taken arsenic, he would not be showing these symptoms.

The first (assuming the conditional premise is probable but less than certain) is good only if we have reason to think not merely that if they had been at home the lights would have been on, but also, that if they were not at home the lights would not be on. If they are the sort of people who always leave the lights on when they go out at night, the lights being off is no evidence that they are not at home—perhaps they have a power cut, or have gone to bed early.

All this, as Adams saw, is just informal reflection on what can be made more precise in a form of Bayes's Theorem:

$$\frac{p_O(H)}{p_O(\neg H)} \times \frac{p_O(E \mid H)}{p_O(E \mid \neg H)} = \frac{p_O(H \mid E)}{p_O(\neg H \mid E)} = \frac{p_N(H)}{p_N(\neg H)}$$

The left-hand equation is a theorem of probability theory, applied to a single probabiity distribution, p_O. 'O' nad 'N' stand for 'old' and 'new' respectively, and represent probabilities prior to learning E, and posterior to learning E. The right-hand equation represents the recommendation that on learning E (and nothing else of relevance) your new probability for H should be your old probability for H given E. Eliminating the middle term, the equation shows how, on learning that E, your new relative values for H and ¬H depend on your old together with these conditional factors: how likely was it that E if H? How likely was it that E if ¬H?

Principles like this are often called principles of 'updating': they tell you how to 'update' your degree of belief in H, from old to new, in the light of new information E. They can, and sometimes do, have that role. But far more prevalent are instances of their use which involve what we might call 'back-dating' ('downdating' doesn't sound quite right). To use them in the updating way, you already have to have foreseen the possibility of the information, E, you receive by perception or testimony; and already, before acquiring it, have a judgement about how likely it is that you will acquire it, under

various hypotheses. This is often extremely unrealistic. We continually see, hear, read in the newspaper, etc., things which we didn't anticipate the possibility of coming across. (I don't mean just outlandish things—I mean perfectly ordinary things that we just hadn't thought about before.) If an observation strikes you as in need of explanation or as the possible basis of an inference relevant to your concerns, you start there, and ask yourself: how likely was it that I *would* get this information, given H? And given ¬H?. Your present 'would haves' record your present opinion about an earlier 'will'.

Thus, anyone in the business of using Bayes's Theorem, without the utterly mythical assumption that we already have degrees of belief about everything, is in the business of evaluating these 'counterfactual probabilities'. (Reconstructing your priors after the event. Note that these priors might concern a time before you were born, or a time before anyone was born: about, say, the survival chances of dinosaurs, given various hypotheses. That is, the priors should *not* be thought of in terms of what you believed yesterday; but of how likely you now think it *was* that this *would* arise, given various hypotheses.) And this sort of judgement should be of interest to any Bayesian, independently of their views about the semantics of 'if'-sentences in English.

4. *Adams's puzzle*.

I shall give it in Adams's words (from p. 129 of *The Logic of Conditionals*).

"Imagine the following situation. We have just entered a room and are standing in front of a metal box with two buttons marked 'A' and 'B' and a light, which is off at the moment, on its front panel. Concerning the light we know the following. It may go on in one minute, and whether it does or not depends on what combinations of buttons A and B, if either, have been pushed a short while before, prior to our entering the room. If exactly one of the buttons has been pushed then the light will go on, but if either both buttons or neither has been pushed then it will stay off. We think it highly unlikely that either button has been pushed, but if either or both were pushed they were pushed independently, the chances of A's having been pushed being 1 in a thousand, while the chance of B's having been pushed is a very remote 1 in a million. In the circumstances we think there is only a very small chance of 1,000,999 in one billion (only very slightly above 1 in a thousand) that the light will go on, but a high probability of 999 in a thousand that *if B was pushed the light will go on*.

Now suppose to our surprise that the light does go on, and consider what we should infer in consequence. Leaving out numerical probabilities for the moment, we would no doubt conclude that the light probably lit because A was pushed and B wasn't, and not because B was pushed and A wasn't (the former being about 1000 times more likely than the latter). Therefore, since A was probably the button pushed, *if B had been pushed the light would not have gone on*, for then both buttons would have been pushed. The point is that the counterfactual would be affirmed [now] despite the fact that the corresponding indicative was very improbable [earlier], because its contrary 'if B was pushed then the light will go on' had a prior probability of 0.999."

(As with any example like this, it may appear artificial, but it is easy to transform it into more realistic cases. For example: John has a rare disease. There are two drugs, D and E,

which would help him. If he takes just one, he'll get better. If he takes both or neither, he'll get worse (though the harmful effect of taking both is not well known, and John won't realise this). Both are in extremely short supply, and it's very unlikely that he'll get either, but it's about 100 times less likely that he'll get E than that he'll get D, and immensely unlikely that he'll get both. (There's no causal connection between the suppliers of the two drugs.) I think 'If John takes E, he'll get better'. Now John does get better. Much the most likely explanation is that he got D. So now I think 'If he had taken E he would have got worse'.)

Adams shows how the informal reasoning above, (leaving aside the counterfactual probability at the end) is vindicated by Bayesian reasoning:

$$\frac{p_N(B)}{p_N(A)} = \frac{p_O(B)}{p_O(A)} \times \frac{p_O(L \mid B)}{p_O(L \mid A)}$$

po(L given B) = po(¬A) = 0.999
po(L given A) = p_O(¬B) = 0.999999
So

$$\frac{p_N(B)}{p_N(A)} = \frac{0.00001}{0.001} \times \frac{0.999}{0.999999} = \frac{999}{999999}$$

And he notes that we did use the fact that the that p_O(L given B) was high, i.e there was a high probability that the light would go on if button B had been pressed.

Yet we now think that it's highly unlikely that if button B had been pressed the light would not have gone on!

Adams then says that he does not see how this last judgement can be a judgement of conditional probability at all! He proves (we won't now go into how, but it's perfectly kosher and the conclusion is pretty obvious) that given the set up:

$p(\neg L \mid B) = p(A \mid B)$

in any probability distribution whatsoever.

Now in our present, new probability distribution, now that we know that the light went on, $p(A \mid B)$ is 0. But we don't think that the probability is 0 that the light wouldn't have gone on if B were pressed.

While in our earlier distribution, before we knew the light went on, we thought it highly likely that if B were pressed, the light would go on! So neither seems to capture our present high probability that if B had been pressed, the light would not have gone on.

I'm going to try to argue later that Adams hasn't exhausted the possibilities, and that there is a way of seeing the latter as a judgement of conditional probability. I'll also be

looking at other 'switching cases' that have been discussed in the literature, though they are a little different from Adams's case. But now I turn to another matter.

5. The 'Law of Total Conditional Probability', or, the battle between two formulas

First consider the unconditional case. You're wondering whether A. You think 'It depends on whether X'. You know (in some cases) how likely it is that A if X, and you know how likely it is that A if ¬X. So how likely is it that X? Given an opinion about that, you can proceed as follows:

$$p(A) = p(A\&X) + p(A\&\neg X) = p(A \mid X).p(X) + p(A \mid \neg X).p(\neg X).$$

This is an instance of what is sometimes called the law of total probability.

For instance you think it's 50-50 whether bag X or bag Y is in front of you. (So Y is in effect ¬X.) Each bag contains 100 balls. A ball is to be selected at random. In bag X (you know) 90% of the balls are red. In bag Y, 10% of the balls are red. How likely is it that you will pick a red ball? (90% × 50%) + (10% × 50%) = 50%. (Of course.)

Now consider the conditional case. You're wondering how likely it is that C if A. It depends on whether X or ¬X. You know the chance of C if A&X; and you know the chance of C if A&¬X. How should you proceed?

First guess: (*) $p(C \mid A) = p(C \mid A\&X).p(X) + p(C \mid A\&\neg X).p(\neg X)$.

For instance, again you think it's 50-50 whether bag X or bag Y is in front of you. (So Y is in effect ¬X.) In bag X, 90% of the red balls have black spots. In bag Y, 10% of the red balls have black spots. How likely is it that if you pick a red ball it will have a black spot? According to the above formula: (90% × 50%) + (10% × 50%) = 50%.

But wait! These bags, X and Y, are the same bags as before. There's a far higher proportion of red balls in bag X than in bag Y. So if I pick a red ball, that makes it likely that it's bag X in front of me, in which case it's likely to have a black spot. Thinking this way, it's well over 50% likely that if I pick a red ball, it will have a black spot.

In fact (*) is not a theorem. Instead we can derive:

(**) $p(C \mid A) = p(C \mid A\&X).p(X \mid A) + p(C \mid A\&\neg X).p(\neg X \mid A)$.

Proof:
$$p(C \mid A) = \frac{p(C\&A)}{p(A)} = \frac{p(C\&A\&X) + p(C\&A\&\neg X)}{p(A)}$$

$$= \frac{p(C \mid A\&X).p(A\&X) + p(C \mid A\&\neg X).p(A\&\neg X)}{p(A)}$$

$$= p(C \mid A\&X).p(X \mid A) + p(C \mid A\&\neg X).p(\neg X \mid A).$$

[(*) and (**) are equal when X is independent of A, but they are not in general equal.]

Applying this to the example, we get (90% × 90%) + (10% × 10%) = 82%
Now, the probability functions may represent credence, or they may represent objective chances. Or they may be a mixture generated by something like what Lewis called the principal principle. In the unconditional case: Suppose I have (keeping things simple) two exclusive and exhaustive hypotheses H_1 and H_2, about the chance of A. Then the probability I should assign to A is $ch(A \mid H_1).p(H_1) + ch(A \mid H_2).p(H_2)$.

In the conditional case, suppose I have two exclusive and exhaustive hypotheses, H_1 and H_2, about the chance of C given A. It might be tempting to think that

$p(C \mid A) = ch(C \mid A\&H_1).p(H_1) + ch(C \mid A\&H_2).p(H_2)$. But that goes with (*) which is not derivable as a theorem. Instead we should have, as the analogue for conditional probabilities:

$p(C \mid A) = ch(C \mid (A\&H_1).p(H_1 \mid A) + ch(C \mid A\&H_2).p(H_2 \mid A)$.

[In fact the failure of (*) is central to Lewis's proof that conditional probabilities are not the probabilities of propositions].

Stefan Kaufmann in a paper in the *Journal of Philosophical Logic* 2004, 'Conditioning against the grain', claims that we often evaluate conditionals by (*) rather than (**) and we should just accept the fact that this is one way of evaluating them, and (**) is another way of evaluating them. He does not report serious empirical work, but says that 9 out of 10 people he asked about a case similar to my balls gave answers in line with (*). He also analyses a few other philosophical puzzles about conditionals that show lines of argument in accordance with (*).

Now (*) is tempting and it isn't a surprise that many people give answers in accordance with it. I have been misled myself on the issue. But the issue is whether these people are making a natural enough mistake, or whether this really is an acceptable way of evaluating conditionals—not as a conditional probability, but as a certain weighted average of conditional probabilities.

That (*) is a mistake has, I think, been shown conclusively in a reply to Kaufmann by Igor Douven (JPL 2008). Douven shows that if we evaluate according to (*), then if we alter the example in a way that should be utterly insignificant and inconsequential, we get a different answer. For instance, go back to the balls.
Bag Y contains 100 balls, 10 of them red, 1 of the red ones with a black spot. Bag Y is wearing thin, so someone puts inside it two bags, Y_1 and Y_2, and distributes the contents between them. Each ball in Y still, as before, has an equal chance of being picked, so this little cosmetic alteration should make no difference. Y_1 and Y_2 each contain 50 balls. Y_1 contains 1 red ball, which has a red spot. Y_2 contains 9 red balls, none with a black spot. Applying (*) before this alteration gave the answer 0.5. Applying (*) after the alteration gives:

$p(B \mid R) = p(B \mid R\&X).p(X) + p(B \mid R\&Y_1).p(Y_1) + p(B \mid R\&Y_2).p(Y_2)$

$= (0.9 \times 0.5) + (1 \times 0.25) + (0 \times 0.25) = 0.7$.

Using instead the kosher principle (**) we get

$p(B \mid R) = p(B \mid R\&X).p(X \mid R) + p(B \mid R\&Y_1).p(Y_1 \mid R) + p(B \mid R\&Y_2).p(Y_2 \mid R)$

$= (0.9 \times 0.9) + (1 \times 0.1) + (0 \times 009) = 0.82$, as before.

And I think this helps us see why (*) uses the wrong weights. The things you want to calculate is something *on the assumption that R*. The probability of X, or Y_1 or Y_2 depends in part on how likely they are in the ¬R-worlds, or possibilities, and that part is irrelevant to how likely something is in the R-worlds, i.e. on the assumption of R.

6. *Back to Adams, and Skyrms*.
Adams used the kosher formula (**) to derive that in any probability distribution, $p(\neg L \mid B) = p(A \mid B)$, and could not find a suitable distribution—neither the posterior nor the prior—for this to be the judgement, after learning that L, that it was very likely that if B had been pressed, the light would not have gone on. So, rather half-heartedly, he suggests, in effect, that we evaluate it according to the non-kosher formula (*). I quote (p. 131):

"We can regard the pushing of button A as putting the electric circuit inside the box into a dispositional state in which pushing B results in the light's not going on, where the mere fact of the light's actually going on can constitute evidence, positive or negative, that A was pushed and the circuit was in that state. ...

A rather simple minded generalization of our prior conditional probability representation which would accommodate counterfactuals entering into the button and light example is as follows. Restrict attention to the counterfactual 'if B would have been pushed the light would not have gone on' whose probability is assumed to be given by $p(B \rightarrow \neg L)$. Generalizing, we may suppose that there are mutually exclusive and exhaustive states $S_1, \ldots S_n$ which are causally independent of B, and which together with B causally determine ¬L. In this case the probability of the counterfactual is plausibly given by

$p(B \rightarrow \neg L) = \sum (i = 1, \ldots n) p_i(S_i) p_0(B\&S_i \rightarrow \neg L)$.

In the particular case under consideration, the two causally independent states are just A and ¬A, and these play the role of dispositional states.

The foregoing 'two factor model' of counterfactual probabilities is admittedly ad hoc. ..... Whereas the counterfactual's probability depends on the *prior* probabilities associated with causal laws, it depends on the *posterior* probabilities of the states. It is the mixture of the prior and posterior probabilities in this combination which accounts for the counterfactual's not satisfying the usual laws of conditional probability."

Skyrms tries to make a virtue of what Adams sees as an unfortunate necessity ('The Prior Propensity Account of Subjunctive Conditionals' in *Ifs* ed. by Harper, Stalnaker and Pearce). He agrees with Adams that indicative conditionals go with epistemic conditional probability. A first shot at counterfactuals might make them go with prior epistemic probabilities, as Adams tried. But on reflection, and because of the kind of counterexample Adams gave, we need to go for the 'prior propensities',

(otherwise known as (objective) chances). In the cases where the chances are known, the two accounts coincide. But if we do not know for certain the values of the prior chances, we may have to do with a weighted average—the expected prior chances. The weights in this average will be epistemic probabilities, and we should use the best ones available for the job—the *posterior* epistemic probabilities. So the 'Basic Assertability Value' of the counterfactual $p \rightarrow q$ is the weighted sum of the various possible hypothetical objective chances of q given p, the weights being your latest epistemic probability for those hypotheses. Thus, he says, from this point of view the model Adams tentatively introduced is not *ad hoc*, but entirely natural.

I will say one thing in favour of Skyrms and a number of things against. In favour, I think we do have a use for both the notion of objective chance and the notion of epistemic probability, in thinking about these matters. It seems to be foreign to Adams's way of thinking to differentiate kinds of probability in this way. But I don't agree with Skyrms that it is just subjunctives/counterfactuals that have some special relation with objective chances. They are just as important to think about, as what we aim at, in indicative conditional judgements about the future like 'If you have the operation you will be cured: the best available opinion is one that matches the conditional chance. Secondly, Adams and Skyrms are using (*) which we have seen to be a highly suspect formula. Go back to the balls in bags. Let's say we were convinced that (**) gave us the right thing to think about how likely it is that you will pick a black spot if you pick a red ball. As a matter of fact, you didn't pick at all, so now we ask how likely it is that you would have got a black spot if you had picked a red ball. There are two hypotheses about the chances: depending on which bag is in front of you. $ch_X(B \mid R) = 0.9$; $ch_Y(B \mid R) = 0.1$. The two hypotheses are equally likely. So on Skyrms's way of looking at things, the answer is 0.5. And this is subject to Douven's objection that the way of assessing is unstable over changes which should be inconsequential.

Relatedly, what Skyrms calls a 'Basis Assertability Value' is, as he knows, not a conditional probability—that's why he gives it a fancy name—so it is quite unclear how it can play the role in inference that, I argued before (following Adams) counterfactuals play. (We simply don't know what if anything we can do with a 'Basic Assertability Value'. And there is another objection coming up in the next section. So Adams's puzzle is still unsolved and we have to see if we can do better.

7. *More Switching Cases*. Adams's puzzle is a switching case: before we learned the light goes on, we think it's likely that if B was pressed, the light will go on. After the light goes on, we think (that's probably because A was pressed so) it's unlikely that if B had been pressed, the light would have gone on.

Other switching cases have been discussed in the literature on counterfactuals. The original was due to Morgenbesser (and they are sometimes called Morgenbesser cases). I'm offered a bet that the coin will land heads. I decline to bet. It is tossed anyway and there is no causal interaction between my declining to bet and it's being tossed. It lands heads. If I had bet on heads I would have won. (Though there was no reason to believe earlier: if I bet on heads I will win.) Here is an example I pursue in a paper 'Counterfactuals and the Benefit of Hindsight'. I miss my plane. As it happens, a later, unpredictable in advance, chance event causes the plane to crash, and all on board are

killed. If I had caught the plane I would be dead. But there was only a minute chance, before I missed the plane, that if I catch the plane I will be killed.

If these judgements are correct, we keep post-antecedent, casually independent of the antecedent, actual facts constant when assessing counterfactuals. This is another argument against Skyrms. Knowing the prior conditional chance doesn't settle matters for the counterfactual: it can be upset by later causally independent facts. I argue that the ultimate value to be assigned to 'If A had been the case, C would have been the case is not necessarily the chance, back then, of C given A, but the chance, back then, of C given A&S where S includes all subsequent facts, casually independent of A, which have some bearing on whether C. It is still a conditional chance, but not just of C given A!

Some people are inclined to reject these switching cases. After all, accepting the bet, or catching the plane, would, as it were, put us in a different possible world where the chances might work out differently. There are two reasons to reject this line of thought, the second more weighty than the first. First, accepting it seems to lead to absurd judgements. For instance, you're watching a lottery draw on television and to your dismay your arch business rival wins a prize—not a big enough prize for him to abandon his business, but big enough for him to put you out of yours. If switching cases were banned, you could say to yourself, 'If I had scratched my nose a minute ago, he very probably would have lost (because then we're in a different possible world where the chances might have worked out differently). What a pity I didn't scratch my nose!'.

Second, and more important, it is these hindsight-ful judgements that feed correctly into our inferential practices. e.g. 'She must have missed her plane', someone says, seeing me later. 'If she had caught that plane she would be dead'.

Here is a more serious inferential example (from my earlier paper). A long time ago, a volcano erupted. It was a slow eruption, the lava creeping slowly forward. At that time, it was very likely that the lava would submerge valley A, and valley B would not be affected—given the lie of the land. However, in the unlikely event of an earthquake of a particular kind at a particular time, the path of the lava would very probably be switched away from valley A, towards valley B, As a matter of fact, that is what happens.

Along comes our geologist, maybe centuries later, making her inference about the eruption. She has already found out about the earthquake. 'That volcano must have erupted', she says, 'for there is lava in valley B and not in valley A, which is what one would expect to find, because of that earthquake.' This could all be formulated in terms of Bayes's Theorem. And *our inferential purposes would not be well served* if we were constrained to use the earlier chances, ignoring the later earthquake. Suppose there was a second volcano whose potential eruption, at the time in question, presented much more danger to valley B, but in the unlikely event of the earthquake, its lava would probably be diverted elsewhere. It's not the earlier chance we want. Only with hindsight is one justified in thinking that if the second volcano had erupted, valley B would not have been submerged.

Adams's switching case differs from these other switching cases in two respects. One, these others are deemed problematic because they depend on a post-antecedent chance event—the coin's landing heads, the plane crash, the earthquake. Adams doesn't tell us, in the original puzzle, anything about the order in which the buttons A and B are pushed. (And while it may make a difference to the metaphysics of the situation, it makes no difference to the epistemology—a point to which I'll return.) If we stipulate that the

pushing (or not) of B is to occur before the pushing (or not) of A, then we have the event in the future of B which affects whether the light goes on. Second, there is an additional element of uncertainty in Adams's example. We are represented as *knowing* that the coin landed heads, the plane crashed, there was an earthquake. We don't *know* that it was A that was pushed, we just now think it is immensely likely. To make these later examples parallel to Adams one would have to consider: I learn that a plane has crashed and have reason to think it was the one I missed. So it's likely that if I had caught my plane I'd be dead. I'm told 'I think I heard them say the coin landed heads'. So I think it's likely that if I had bet on heads I would have won. There's strong evidence that there was an earthquake; so it's likely that if that volcano had erupted, valley B would be submerged. Then we have parity between the two kinds of example.

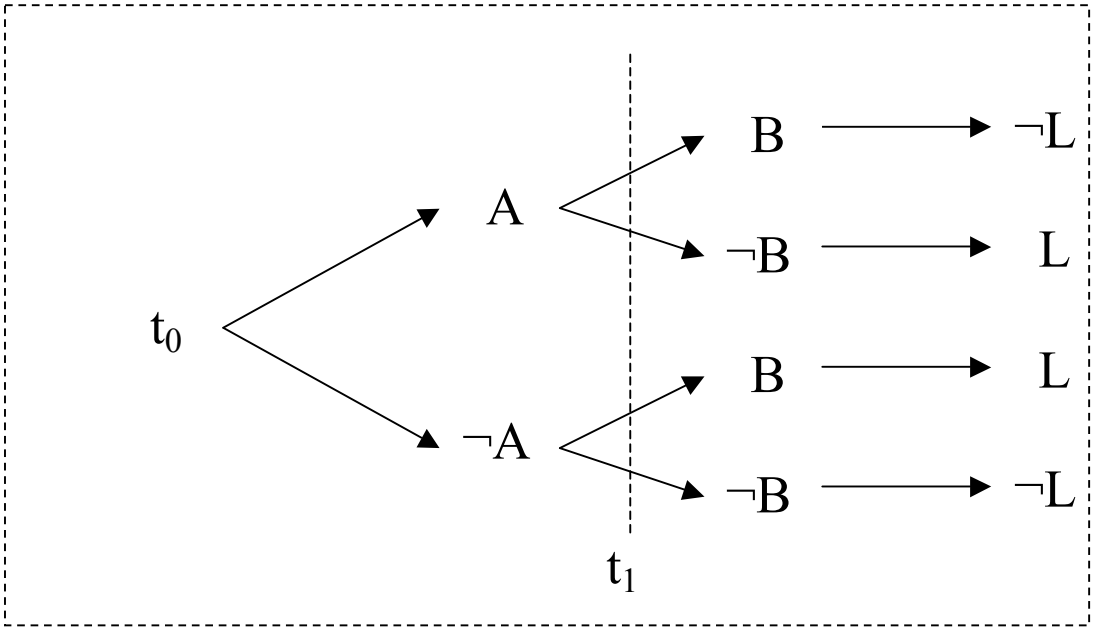
8. *My attempt at a solution to Adams's puzzle*.

If our approach to the later switching puzzles is right, we do *not* always, when assessing a counterfactual, want to estimate the chance, back then, of C given A, but of C given A *and* any subsequent facts causally independent of A which have a bearing on whether C. The trouble is, in the uncertain shifting cases, we're not sure what the facts are, i.e. we're not sure which conditional probability we're aiming at: the chance of death given that I catch the plane and it crashes, or the chance of death given that I catch the plane and it doesn't crash. In Adams's example, the chance that the light goes on given that B is pushed and A is pushed; or the chances the light goes on given that B is pushed and A is not pushed. If A was pushed, we need $p(L \mid A\&B)$. If A was not pushed, we need $p(L \mid \neg A\&B)$. It's hard to deny that these two just be weighted by the probability of A and of ¬A. But is the resulting weighted average a conditional probability? Or are we using the bad old (*)?

Here's another line of thought that turns out to be more helpful. As I said, Adams doesn't tell us about the order (if any) in which the buttons are to be pushed (or not pushed), and he doesn't have to, because epistemologically it makes no difference. And it occurred to me that the same is true of the other switching cases. It makes no difference epistemologically to what I think, whether it was a late chance event that brought down the plane, or whether there was a bomb on board planted yesterday. It makes no difference to my thinking 'If I had bet on heads I would have won', whether it was an ordinary coin that landed heads, or whether it was a double-headed coin that landed heads.

So, without loss of generality from the epistemological point of view, we can model Adams's case with the pushing (or not) of A coming first, the pushing (or not) of B later. That way we finesse the problems of keeping constant chance facts subsequent to the antecedent.

The picture is like this:



Adams showed that in thinking about 'if B had been pushed the light would not have gone on' we don't want our present $p(\neg L \mid B)$, which is 0; nor do we want that associated with the starting point t_o. But Adams missed this possibility: the appropriate past hypothetical standpoint for evaluating 'If B had been pushed the light would not have gone on', is of someone at t_1. A is already settled and they think it's around 99.9% likely that A. B is still a chance event in the future, independent of A. They know that $ch(\neg L \mid B) = 1$ if A, and 0 if $\neg A$. Because of the independence they can us the 'bad' formula: $p(\neg L \mid B) = p(\neg L \mid B\&A).p(A) + p(\neg L \mid B\&\neg A).p(\neg A) \approx (1 \times 0.999) + (0 \times 0.001) \approx 0.999$.

I didn't know of Adams early work, and didn't know of Adams until, one day in late 1975 or early 1976, David Hamlyn, my head of department and Editor of *Mind*, came into my office with *The Logic of Conditionals* and asked if I would review it for *Mind*. I had a prior interest in probability and a prior interest in conditionals but I had not connected the two. Adams convinced me, and that changed the course of my philosophical life.

In that review I wrote near the beginning 'What emerges is a most interesting theory of the indicative conditional; best of all, one in terms of which the notorious counterfactual conditional can easily be explained'. I would now wish to retract the word 'easily'. But I still believe that Adams's theory of counterfactuals was on the right lines.