Triviality Pursuit[1]

Alan Hájek

1. Setting the scene

Like Jason of the *Friday the 13th* franchise, the thesis that probabilities of conditionals are conditional probabilities has enjoyed a number of resurrections. Lewis (1976) appeared to have killed it in his famous triviality results; but van Fraassen (1976) resuscitated it. Stalnaker (1976) attacked it some more; but Rehder (1982) revived it. Further triviality results appeared to deliver fatal blows to it (e.g., Lewis 1986, Hájek 1989, Hájek 1994, Hall 1994, Milne 2003); yet it lives on still, albeit transmogrified, in the writings of Edgington (1995) and Bennett (2003). In this paper I will assault it yet again; but it will doubtless survive in some form in further sequels.

While I think that the various negative and positive results against the thesis are interesting in their own right—and I will offer a negative result of my own—I don't want this paper to be merely an exercise in theorem-proving. I will begin by motivating the thesis and discussing some of the philosophical ramifications of its fluctuating fortunes. I will canvas various reasons, old and new, why the thesis seems plausible, and why we should care about its fate. There are good reasons why it keeps making comebacks. I will look at some objections to Lewis's famous results, and thus some reasons for the pursuit of further triviality results. I will generalize Lewis's results in ways that meet the objections. I will conclude with some reflections on the demise of the thesis—or otherwise.

2. The thesis—or theses

"The probability that Collingwood will win if they are ahead at half time is high." This seems to say that the probability of a certain conditional is high. And it seems to say exactly the same thing as: "The probability that Collingwood will win given that they are ahead at half time is high." Moreover, the probability of the conditional seems to rise or fall in lockstep with the conditional probability. Symbolically:

$$P(\text{ahead} \rightarrow \text{win}) = P(\text{win} \mid \text{ahead}),$$

where '$\rightarrow$' represents the conditional 'if … then', and $P(\text{win} \mid \text{ahead})$ is given by the usual ratio formula for conditional probability, $P(\text{win \& ahead})/P(\text{ahead})$. But this equation seems to generalize to all sentences for which the conditional probability is defined:

(PCCP) $\qquad P(A \rightarrow B) = P(B \mid A)$

for all A and B in the domain of P, provided $P(A) > 0$.

And offhand, it seems to generalize to all probability functions:

For all probability functions P, PCCP.

However, *this* thesis is easily refuted. Consider a 3-ticket lottery, and let L_i = ticket i wins, i = 1, 2, 3. Model it with the obvious 3-world probability space, with $P(L_i) = 1/3$ for all i. Then the conditional probability $P(L_1 \mid L_1 \text{ v } L_2)$ is obviously ½. But all unconditional probabilities are a multiple of 1/3, so in particular the unconditional probability $P((L_1 \text{ v } L_2) \rightarrow L_1)$ is a multiple of 1/3. And ½ is not a multiple of 1/3.

Clearly, then, we need to restrict suitably the probability functions for which PCCP is supposed to hold. For each such restriction, we get a corresponding thesis that probabilities (so restricted) of conditionals are conditional probabilities (so restricted). For example, Stalnaker (1970) thought that PCCP holds for all probability functions that could represent a rational agent's systems of beliefs—all *rational credence* functions. This has come to be

known in the literature as *Stalnaker's Thesis*. I adopt that moniker with some misgivings since, to be fair to Stalnaker, some years later he went on to renounce the Thesis, and he even proved a striking triviality result against it (in Stalnaker 1976). And he was the original architect of causal decision theory as a rival to Jeffrey's evidential decision theory, with probabilities of conditionals replacing Jeffrey's conditional probabilities (more on this shortly). In order to be a rival to, rather than merely a restatement of, Jeffrey's theory, probabilities of conditionals had better come apart from conditional probabilities for rational agents at least sometimes. So this later important project of Stalnaker's *requires* Stalnaker's Thesis to be false! That said, he deserves much credit for inspiring the fertile literature that has grown out of his original Thesis, and at least the moniker serves to remind us of that.

We could also come to the restriction on the probability functions for which PCCP allegedly holds from the other side, so to speak. I have just provided a negative result—a very modest one—that *some* restriction is required. Further negative results will impose further restrictions—the stronger the restrictions, the more interesting the result. We can imagine various theses of various strengths, which might occupy intermediate ground: weaker than Stalnaker's Thesis, and indeed weak enough to evade all the triviality results, but arguably strong enough still to be of interest. Hájek and Hall (1994) survey some candidates for such theses.

3. Why care about the theses?
<u>The semantics of conditionals</u>

The larger the class of probability functions for which PCCP holds, the better for allies of Stalnaker's Thesis. Stalnaker argued that PCCP's holding for all rational credence functions had ramifications for the semantics of conditionals. Reading the equation from left to right, we can regard it as constraining the

assignment of a probability to a conditional. This, in turn, constrains the logic of the conditional. As Stalnaker (1970) observes,

> although the interpretation of probability is controversial, the abstract calculus is a relatively well defined and well established mathematical theory. In contrast to this, there is little agreement about the logic of conditional sentences... Probability theory could be a source of insight into the formal structure of conditional sentences" (p. 107[2]).

In particular, PCCP implies (a probabilistic version of) conditional excluded middle. This was the locus of disagreement between Stalnaker and Lewis on the logic of counterfactuals, the former advocating it, the latter rejecting it. The intuitiveness of PCCP seemed to provide Stalnaker with a weighty argument in his favor. No surprise, then, that Lewis was motivated to expose PCCP's limited tenability.

Probabilistic validity

Meanwhile, Adams (1975) was developing an important approach to conditionals of his own. He believed that conditionals do not have truth conditions, and thus that the traditional 'truth-preservation' account of validity of arguments was inadequate for ones in which conditionals appears. But he was happy to speak of probabilities attaching to conditionals, and he offered a probabilistic surrogate for validity. Define the *uncertainty* of X to be $1 - P(X)$. A *probabilistically valid argument* is one in which necessarily the sum of the uncertainties of the premises is at least as great as that of the conclusion. Adams invoked PCCP to govern the assignment of probabilities to indicative conditionals, arguing that the resulting scheme respected intuitions about which inferences were reasonable, and which not.

While Adams appealed to the same thesis (at least superficially) as Stalnaker did, unlike Stalnaker he did not regard the probability of a conditional as

---

[2] Page references are to Harper et al. (1981) for all articles that are cited as appearing there.

probability of its *truth*. Further, Adams' 'probabilities' of conditionals do not conform to the usual probability calculus. In particular, according to him conditionals are not fit to enter into Boolean combinations with other sentences, and so the domain of a probability function that allows conditionals to be among its arguments is not a field. Thus, Lewis (1976) suggested that they be called "assertabilities" instead, a practice that has been widely adopted subsequently. So the left hand side of PCCP is perhaps best read as "the assertability of *B*, if *A*" on the Adams view. This conditional assertability then goes by $P(B \mid A)$. This is the famous *Adams Thesis*, a touchstone of much of the literature on conditionals.

Implicit definition of the conditional

Still reading PCCP from left to right, we might even regard it as providing an implicit *definition* of the conditional: it is whatever connective obeys the equation. Now, it might seem odd that a logical operator should be defined probabilistically. It seems less odd if we regard it as encoding a rational agent's updating dispositions, an idea that we find in Ramsey (1965) and developed further by Levi (1980).

Iterated conditionals

PCCP would also illuminate right-nested iterated conditionals. It implies:

$$P(A \rightarrow (B \rightarrow C)) = P(B \rightarrow C \mid A)$$

(I assume here that the iterated conditional is in the domain of P.) Now define $P_A(\_) = P(\_ \mid A)$, the result of conditioning P on A. Then we may continue:

$$P(B \rightarrow C \mid A) = P_A(B \rightarrow C)$$
$$= P_A(C \mid B), \text{ by PCCP applied to } P_A.$$

The eye-glazing probability of the iterated conditional is elegantly simplified to a perfectly manageable conditional probability.

The interpretation of conditional probability

Reading PCCP from right to left, we enhance our understanding of conditional probability. de Finetti (1972) complained that the usual ratio gives the formula, but not the meaning, of conditional probability. PCCP would characterize it: it is simply the probability of a conditional! So the rather technical notion of conditional probability would be reduced to unconditional probability, which is arguably a simpler notion, and 'if … then', which is part of ordinary English.

Conditional probability as a special case of unconditional probability

Pursuing this thought further, Kolmogorov already offered a reduction of conditional probability to unconditional probabilities—namely, to a ratio of them—with the formula for $P(B \mid A)$ that is now standard. PCCP promises a more thoroughgoing reduction: conditional probability just *is* unconditional probability, with a particular content (that of a conditional).

The Judy Benjamin problem

Again exploiting the right-to-left reading of PCCP, it would allow us to solve an outstanding problem for Bayesian epistemology, what van Fraassen (1989) calls "the Judy Benjamin problem".[3] The general problem for probability kinematics is: given a prior probability function P, and the imposition of some constraint on the posterior probability function, what should this posterior be? This problem has a unique solution for certain constraints—for example:

1. Assign probability 1 to some proposition[4] E, while preserving the odds of all propositions that imply E. Solution: conditionalize P on E.

---

[3] I owe this point to Ned Hall.
[4] I will speak interchangeably of 'propositions' and 'sentences', and I will use both set-theoretic operations and sentential connectives, in a way that I hope is perspicuous.

2. Assign probabilities $p_1, ... p_n$ to the cells of the partition $\{E_1, ... , E_n\}$, while preserving the odds of all propositions within each cell. Solution: Jeffrey conditionalize P on this partition, according to the specification.

But consider the constraint:

3. Assign conditional probability p to B, given A.

The Judy Benjamin problem is that of finding a rule for transforming a prior, subject to this third constraint.

van Fraassen provides arguments for three distinct such rules, and suggests that such uniqueness results "will not extend to more broadly applicable rules in general probability kinematics. In that case rationality will not dictate epistemic procedure even when we decide that it shall be rule governed" (1989, p. 343). But if Stalnaker's Thesis were true, a particularly simple solution would present itself. After all, constraint 3 would then be equivalent to:

3'. Assign probability p to $A \rightarrow B$,

and this is uniquely met by a simple Jeffrey conditioning on the partition $\{A \rightarrow B, \neg(A \rightarrow B)\}$, assuming that the odds of propositions within each cell are to remain the same.

A new constraint on unconditional probabilities

Now let's read PCCP symmetrically, not privileging either direction. We can regard it simply as a new constraint on three unconditional probabilities: P($A \rightarrow B$), P(A & B), and $P(A)$. Understood this way, it plays a role akin to that of the additivity axiom—it is part of probability theory. The Bayesian, in turn, would then presumably want to appropriate it as a coherence constraint, one obeyed by any ideally rational agent.

        *   *   *   *   *

Such is the bounty that PCCP promises. If Stalnaker's Thesis were tenable, logic, probability theory, and Bayesian epistemology would all be enriched. But its untenability is also philosophically fecund.

<u>Decision theory</u>

Here is a quick recap of the lore regarding decision theory. In the beginning there were formulations of Bayesian decision theory by Ramsey (1931/1990) and Savage (1954); these were refined and presented in a philosopher-friendly manner by Jeffrey (1966), and his evidential decision theory became the philosophical orthodoxy.

Then along came Newcomb's problem. Evidential decision theory apparently recommended one-boxing, while many regarded this as irrational. So causal decision theory, which recommended two-boxing, was born. Indeed, a number of versions of it were developed—first by Stalnaker (1976), whose theory was elaborated by Gibbard and Harper (1978); then by authors such as Skyrms (1980), Lewis (1981), Sobel (1994), and Joyce (1999). But—this is essential to the lore— these versions are merely stylistic variants of one another. Lewis (1981) writes: "We causal decision theorists share one common idea, and differ mainly on matters of emphasis and formulation" (5). Or Harper and Skyrms (1988): "It can be argued … that the various forms of causal decision theory are equivalent – that an adequate version of any one of these approaches will be interdefinable with adequate versions of the others." (Introduction, x.) Joyce (1999) concurs: "I think this is basically right; … the causal decision theorist can adopt an attitude of benign indifference …" (171-2). We can thus speak of 'causal decision theor**y**' in the singular.

Finally, and crucially, the lore insists that evidential decision theory and causal decision theory typically agree. In garden-variety decision problems, such as whether you should carry your umbrella to work, or whether you should bring white or red wine to the dinner party, the two theories give the same advice. It is only in rather recherché Newcomb-like cases that the theories come apart.

I have long been puzzled by the lore. For starters, I am not convinced that all the various versions of causal decision theory *are* equivalent. Indeed, surely some

pairs of them are *not* equivalent. For example, while Gibbard and Harper (1979) weight their utilities with probabilities of *counterfactuals*, Skyrms (1980) weights his with expectations of *conditional chances*. Gibbard and Harper's counterfactuals invoke other possible worlds, while Skyrms's chances are completely anchored in the *actual* world. When the counterfactuals concern suppositions about the chances *being different from what they actually are*, presumably these two versions will come apart. Moreover, where Gibbard and Harper's formulation involves *unconditional* probabilities (of counterfactuals), Skyrms's involves *conditional* probabilities (conditional chances). Here we enter triviality result territory, where the different behaviors of these quantities are brought into relief—more on that shortly. And an agent's credences may outrun her expectations of conditional chances, since she may assign credences to propositions that she thinks *lack* chances. For example, de Finetti (1972) had plenty of credences, but he thought that chance statements were nonsense, so his expectations of them were either 0 (nonsense being known not to be true) or undefined (nonsense being unfit for credence assignments at all). Or someone can have plenty of credences, while thinking like Cartwright (1999) that only rather special physical systems have chances, and thus expectations of conditional chances. And even those of us who think there are *lots* of chances out there, may still think there are *some* chance gaps, which may nevertheless be proper objects of credences; again we will find credences where there are no corresponding expectations of conditional chances. If such gaps are among the options of a decision problem for an agent, the Gibbard-Harper formulation and the Skyrms formulation will presumably come apart, *contra* the lore.

This, in turn, makes me suspicious of the other centerpiece of the lore: that causal decision theory and evidential decision theory typically agree. Certainly, if the causal decision theories don't all agree with each other, even on garden variety decision problems, then they don't all agree with evidential decision theory on those problems! Take Gibbard and Harper's version as representative of

causal decision theory—according to the lore, it should do as well as any other version. It replaces evidential decision theory's *conditional probabilities* of the form P(state | act) with *probabilities of conditionals* of the form P(act $\Box\!\!\rightarrow$ state). But apparently the triviality results collectively teach us that probabilities of conditionals come apart from conditional probabilities easily, often, and by a lot.[5] (My result in this paper will only drive this point home further.) And they do so for 'umbrella' and 'wine' propositions as much as for 'Newcomb' propositions. To be sure, expected utilities are sums of products of such probabilities and utilities, so they might somehow compensate for the differences between the probabilities of conditionals and the corresponding conditional probabilities taken individually. Well, fingers crossed! It would seem almost miraculous if all the differences between the two kinds of quantities washed out like this. And this article of faith is surely not something that we should be left to intuit—it requires proof. The foundations of decision theory deserve no less.

Turning this thought on its head: causal decision theorists should hope that the various formulations are *not* equivalent. If they were, we should have no confidence that causal decision theory delivers correct results in garden-variety decision problems—for we *are* confident that evidential decision theory does. After all, if the Gibbard-Harper formulation comes apart from evidential decision theory even for those problems, then it is in trouble; and if it is a good representative of all the causal decision theories, then they all are in trouble. And yet ironically, the causal decision theorist presumably wants the triviality results to have *some* purchase. After all, if probabilities of conditionals really are conditional probabilities, then the Gibbard-Harper theory is equivalent to Jeffrey's; but since the latter is simpler, then the former is otiose, as are all its allegedly equivalent reformulations. Thirty-five years of literature on causal

---

[5] At the end of the paper I will briefly discuss 'no truth value' accounts of indicative conditionals, which seemingly evade the triviality results, suggesting that even these accounts may not evade them altogether after all.

decision theory would then appear to be a big red herring. So the causal decision theorist who upholds the lore should hope that the triviality results work—but not *too* well!

4. Why believe the theses?

So much for reasons for caring about Stalnaker's Thesis and other theses with various restrictions on the probability functions that conform to PCCP. Why should we believe them? Consider Stalnaker's Thesis in particular—hereafter, 'the Thesis'.

It sounds right

We have already seen what is surely the main reason: *it sounds right*. (See van Fraassen 1976, Bennett 2003.) The slogan that 'probabilities of conditionals are conditional probabilities' rolls easily off the tongue. And case-by-case evidence seems to support it—remember the 'Collingwood' example.

I think this is a rather weak reason to believe the Thesis. The slogan that 'probabilities of limits are limits of probabilities' rolls off the tongue with comparable ease, and we could marshal much case-by-case evidence to support it. But *that* thesis is controversial—it is guaranteed to hold only for countably additive probability functions, and friends of merely finite additivity such as de Finetti will insist that it does not generalize. The methodology of case-by-case checking is unreliable in any case. I surmise that we typically test mentally a handful of simple cases like the 'Collingwood' one, give them a tick, and then hastily generalize our positive appraisal of them to the infinitely many remaining cases. But we know that this methodology *over*-generalizes. Exhibit A: the claim that in the probability space of the 3-ticket lottery, the probability of 'if one of the first two tickets wins, then the first one wins' equals the conditional probability of 'the first one wins, given one of the first two tickets wins' *sounds right*. But we know that it is wrong, as we saw above. More generally, we apparently know that

probabilities of conditionals come apart from conditional probabilities easily, often, and by a lot. But they keep *sounding* like they should agree.

The two sides are structurally alike

A better approach is to argue more systematically why the two sides of PCCP should coincide. After all, they seem *structurally* and *semantically* alike. On both sides, the probability takes wide scope. Both sides involve a restriction to the cases in which *A* is true; the probability is to measure, among those cases, how likely are the *B* cases.

The Ramsey test

In the most famous passage ever written on conditionals, Ramsey (1965) writes: "If two people are arguing 'If *p* will *q*?' and are both in doubt as to *p*, they are hypothetically adding *p* to their stock of knowledge and arguing on that basis about *q*." Applied to your own evaluation of 'if *A*, then *B*', the Ramsey test bids you first to hypothetically add *A* to your system of beliefs, minimally revising what you currently believe in order to do so; second, to evaluate *B* on the basis of your revised body of beliefs. $P(A \rightarrow B)$ measures how well the conditional performs on Ramsey's test. But apparently $P(B \mid A)$ does too. For conditioning on *A* prima facie seems to capture the notion of 'minimally revising what you currently believe in order to accommodate *A*'; and your evaluation of *B* in your new belief state $P(\_ \mid A)$ is just $P(B \mid A)$.

Better still, Ramsey offers a version of the test based on the corresponding conditional probability. He goes on to say: "We can say that they are fixing their degrees of belief in *q*, given *p*." Now the identification of the probability of the conditional with the corresponding conditional probability is even more explicit. The Ramsey test has much currency even today—so much so that it is often taken to be the starting point of theorizing about conditionals.

Ramsey's imaginary protagonists are arguing about an *indicative* conditional, and plausibly the Ramsey test should be restricted to such conditionals. But this is all to the good as far as the Thesis is concerned. For conditional probability has a similarly 'indicative', rather than 'counterfactual', nature. Correspondingly, conditioning is meant to capture updating rather than supposing. We assume that A is compatible with your belief state, and the indicative conditional $A \rightarrow B$ has you reason about B under the assumption that A *is* the case. But so does the conditional probability, $P(B \mid A)$.

Adams' Thesis

Assertability is said to go by subjective probability of truth. (Or at least it usually does, though not always: for example, the assertability of 'A but B' differs from that of 'A and B', yet the two have the same subjective probability, namely $P(A \ \& \ B)$.) At first sight, the indicative conditional appears to provide a counterexample to this dictum. After all, according to the popular understanding of Adams' Thesis, the assertability of $A \rightarrow B$ equals $P(B \mid A)$—and even someone who, unlike Adams, believes that the indicative conditional has truth conditions, may find Adams' Thesis about its assertability compelling. (Of course, someone *like* Adams thinks that the indicative conditional provides a counterexample to the dictum, since according to him, there is no such thing as the probability of its *truth*.) It would be nice for proponents of the truth-conditional view of the indicative conditional if it conformed to the dictum. If Stalnaker's Thesis were true, it would then explain Adams' Thesis admirably: the assertability of $A \rightarrow B$ goes by $P(A \rightarrow B)$ (as per the dictum), which in turn equals $P(B \mid A)$. We infer the truth of Stalnaker's Thesis from Adams' Thesis since the Thesis provides a good—perhaps the best—explanation of it. This is the motivation for the thesis discussed in Lewis (1976) and (1986).

     *   *   *   *   *   *

So much for reasons to want Stalnaker's Thesis to be true, and to believe that it *is* true. However, there are much stronger reasons to believe that it is false. Lewis initiated the industry of providing them in his classic (1976) paper, with his so-called 'triviality results'. To these I now turn.

5. Lewis's triviality results

While Lewis presents two triviality results in this paper, logically speaking there is only one: the second one entails the first, and their proofs are almost identical—so much so that it is perhaps a little surprising that Lewis distinguishes them. So let us confine our attention to the stronger result. A number of authors misstate it. For example, Adams (1975) claims that in his (second) triviality result, "Lewis assumes that *probabilities change by conditionalization"* (8). But Lewis' *result* makes no such assumption. To be sure, this assumption plays an important role in motivating the *interest* of the result; but that's another matter (and even the interest of the result is not so-confined, as I will soon argue). I mention this because, as we will see, Lewis's result is not exactly intuitive, and it is rather complicated to state precisely. It is surprisingly hard both to paraphrase it and to get it exactly right. (Exercise for those who think they already know the result: try doing so yourself!)

Some notation and terminology: $P_X(\_)$ is the function that one gets from $P(\_)$ by conditioning on X—that is, for all A and X in the domain of P,

$P_X(A) = P(A \mid X)$, provided $P(X) > 0$.

If PCCP holds for the pair $<\rightarrow, P>$, we will say that $\rightarrow$ is a *probability conditional for P*. If $\rightarrow$ is a probability conditional for every probability function in some class of probability functions, we will call $\rightarrow$ a *probability conditional for* the class. Say that P is a *trivial* probability function if there are not two propositions A and B such that $P(A) < 1$, $P(A \& B) > 0$ and $P(A \& \neg B) > 0$.

Lewis assumes the following, but he only sketches its proof. I will give the proof in more detail.

Import-Export Lemma:

If PCCP holds throughout a class of probability functions that is closed under conditioning, then the following also holds throughout the class:

$P(A \rightarrow B \mid C) = P(B \mid A \ \& \ C)$, if $P(A \ \& \ C) > 0$.

Proof:

Suppose that PCCP holds throughout a class $\mathscr{C}$ of probability functions that is closed under conditioning, let P belong to $\mathscr{C}$, and suppose that $P(A \ \& \ C) > 0$.

$P(A \rightarrow B \mid C) = P_C(A \rightarrow B)$, by definition of P_C

$= P_C(B \mid A)$, by PCCP applied to P_C,

since $P(A \ \& \ C) > 0$ and hence $P_C(A) > 0$

$= P_C(A \ \& \ B)/P_C(A)$,

by the formula for conditional probability, since $P_C(A) > 0$

$= P(A \ \& \ B \mid C)/P(A \mid C)$, by definition of P_C

$= P((A \ \& \ B) \ \& \ C)/P(A \ \& \ C)$, by the formula for conditional probability, since $P(A \ \& \ C) > 0$ and hence $P(C) > 0$

$= P(B \mid A \ \& \ C)$, by the formula for conditional probability.

So what exactly *is* Lewis's result? Here it is.

Theorem:

If a class of probability functions is closed under conditioning, then there is no probability conditional for that class unless the class consists entirely of trivial probability functions.

Proof:

Suppose for reductio that PCCP holds throughout a class of probability functions that is closed under conditioning. Take any probability function P in the class and any sentences A and B such that $P(A \ \& \ B) > 0$ and $P(A \ \& \ \neg B) > 0$.

$P(A \rightarrow B) = P(A \rightarrow B \mid B)P(B) + P(A \rightarrow B \mid \neg B)P(\neg B)$

> by the law of total probability, since $P(A \ \& \ B)$ and hence $P(B) > 0$,
>
> and since $P(A \ \& \ \neg B)$ and hence $P(\neg B) > 0$.
>
> $= P(B \mid A \ \& \ B)P(B) + P(B \mid A \ \& \ \neg B)P(\neg B)$, by the Import-Export Lemma
>
> $= P(B) + 0$
>
> $= P(B)$.

6. Questioning Lewis's results

Against conditioning

Conditioning has a central place in Bayesian orthodoxy. However, for some purposes it seems inadequate—for example, for updating credences on sentences involving indexicals (especially temporal ones), or for modeling certainty loss. (See Titelbaum 2008.) Indeed, some authors have questioned the rationality of updating by conditioning quite generally. (See Bacchus, Kyburg and Thalos 1990, Hild 1998.) Let us grant for the sake of argument, *pace* Lewis, that conditioning is *not* a rational updating method. This is not sufficient to undermine the importance of Lewis's result, however, for it does not impugn the rationality of two *different* agents whose probability functions are related by conditioning. (Note that according to Adams' presentation of the result, which I quoted above, conditioning *not* being a rational updating method *is* sufficient to undermine the importance of the result—this shows how getting the details right matters.) Still, it would be nice not to make *any* assumption at all about $\rightarrow$ being a probability conditional for functions related by conditioning.

While van Fraassen is not opposed to conditioning—his permissive epistemology is hard to offend—he is hardly its most ardent supporter. After all, he believes that rationality does not require conditioning. (See his 1989.) Indeed, by his lights rationality does not require rule-following at all, permitting wholesale discarding of one's priors, and epistemic leaps that are not grounded solely in evidence. While not undermining Lewis's result itself, this view

threatens to undercut its importance a little. Conditioning no longer enjoys pride of place. So we should welcome a result that obviates the need to appeal to it.

The conditional as probability-sensitive

Lewis assumes that the conditional is fixed, etched in stone irrespective of the probability function in whose scope it appears. In particular, he assumes that the same '$\rightarrow$' appears in the arguments of P, P_B, and $P_{\neg B}$. But van Fraassen (1976) demurs: "Would it not seem rather, that our probabilities are inextricably involved in the way we represent the possibilities, and nearness relations among them, to ourselves ... if our ideas about the one change, will we not revise our modelling of the other?" (274) He suggests that the conditional may be radically sensitive to the probability function—every change in the latter brings about a change in the former. On this view, we do not have a single conditional at all, but rather one conditional that we could denote '$\rightarrow_P$' when it appears in the scope of P, another conditional '$\rightarrow_{P_B}$' when it appears in the scope of P_B, and yet another conditional '$\rightarrow_{P_{\neg B}}$' when it appears in the scope of $P_{\neg B}$.[6]

7. Generalizing Lewis's triviality result

Now I will generalize Lewis's triviality result. Doing so will go a long way towards answering these objections, and it will have other benefits besides.

For our purposes, a probability *revision rule* takes as input an initial probability distribution and a proposition, and yields as output a (new) probability distribution. We will regard an agent's system of beliefs as being represented by a probability function, and a change in the belief system to accommodate certainty

---

[6] One wonders what the conditional becomes when it appears in the scope of *two or more* probability functions—when, for example, P assigns probabilities to various hypotheses about another function P^*'s assignments! (Shades here of the contextualist and relativist literature on reported speech and eavesdropping cases.)

about the truth of some proposition as taking place according to some revision rule, such that the proposition gets probability 1 after the revision.

As we have seen, Lewis focuses on the revision rule of conditioning. However, there are other well-known rules—here are four.

- *Imaging* on some proposition E is the rule that moves all the probability from each $\neg E$ world to its nearest neighbor inside E—'nearest' as determined by the relevant similarity relation. (This assumes, as Stalnaker (1968) does, that there is always a unique such neighbor.)

- A *blurred imaging* on E removes all the probability from a $\neg E$ world, but spreads it over more than one 'nearest' E world. Various ways of doing this have been proposed—see Gärdenfors (1982) and Lewis (1986, p. 318) for more details and references.

- *Maxent* is the rule that takes an initial probability function and a constraint, and yields the probability function of *maximum entropy* that satisfies the constraint. It will suffice for my purposes to define 'entropy' for discrete cases: if countably many worlds $i = 1, 2, \ldots$ receive positive probabilities $P(i)$ then the entropy of this distribution is $-\sum_i P(i) \log P(i)$. If the constraint on possible final probability measures is 'the probability of E equals 1', then we have the case of accommodating E by maxent; the distribution that is produced is uniform over E.

- *Minxent*, or *Minimum Cross Entropy*, revises a probability function P to the *nearest* probability function Q that meets a given constraint, where 'nearness' is measured by the so-called Kullback-Leibler distance, which for discrete distributions P and Q is given by:

$$D_{KL}(P \parallel Q) = \sum_i P(i) \log \boxed{\times}$$

I will show that similar results to Lewis' are available for imaging, various blurred imagings, maxent, minxent, and other revision rules. Lewis's second result will fall out as a corollary, but my result will be more general than it is.

Like his result, and unlike those of Stalnaker (1976) and Hall (1994), mine makes no assumption about the logic of the conditional.

8. Boldness, conservativeness, moderation

Suppose we have an initial probability distribution P, and we want to revise it in order to accommodate a proposition E. We use some revision rule to derive a distribution P_E. (Here I use more generally the notation that previously I used just for the revision rule of conditioning.) Call the rule *bold* if for any P and for any E such that $P(E) > 0$, $P_E(E) = 1$. 'Bold', for two reasons. Firstly, the rule is prepared to take as input *any* initial probability function and proposition. Secondly, the function produced as output by the rule is not afraid to commit itself fully, giving probability 1 to the proposition (as it must, in order for the proposition to be genuinely accommodated). Bold revision rules take us from some initial probability distribution to a new distribution that is fully concentrated on any proposition that we specify. Conditioning, imaging, the varieties of blurred imaging, maxent and minxent are all bold.

Furthermore, some revision rules are what we may call *conservative.* Such a rule takes a function P and a proposition E to a function P' with the following property: for any A that implies E, $P'(A) \geq P(A)$. A conservative rule never decreases the probabilities of propositions that imply the proposition that is accommodated. Conditioning is conservative; so is imaging and various blurred imagings, assuming that the similarity relation obeys centering (each world is closer to itself than any other world is to it). Still more rules are what I will call *moderate:* for any A that implies E, if $P(A) > 0$, then $P'(A) > 0$. Moderate rules can decrease the probabilities of propositions that imply the proposition that is accommodated, but never all the way to 0.

Moderation is surely a desideratum of a revision rule: your belief system would be highly fragile if you could suddenly *fully* disbelieve some proposition A that previously you gave some credence, when you learned something implied by

A. By the lights of your original credence function, *A* is confirmed (assuming the usual Bayesian account of confirmation); but by the lights of your new credence function, it is maximally disconfirmed. A little more precisely: if $P(E) < 1$, *A* entails *E*, and $P(A) > 0$, then *A* and *E* are initially correlated:

$P(E \mid A) > P(E)$.

But if *P* is non-trivial and is revised by a rule that it is not moderate, then there are some *A* and *E* such that all these conditions are met, *E* is the strongest proposition that is learned, and yet *A* and *E* are *maximally anti-correlated* after the revision:

$P(E \mid A) = 0$.

It is hard to see how anything that bears *so radically* on the correlation between *E* and *A* has been learned. The radical revision in the assessment of *A*'s evidential bearing on *E* seems entirely gratuitous, and certainly not justified by the evidence *E* itself.

All conservative rules are moderate. Moreover, so is maxent and minxent (at least for the finite cases that I have defined above). So too are various blurred imagings, even when the similarity relation obeys merely weak centering: any world is one of the nearest worlds to itself.

Even if, *pace* Lewis, probability functions in a rational epistemic history are not related by conditioning, a negative result analogous to his holds, provided that they are related by *some* particular rule that is bold and moderate; and similarly if the probability functions of two different rational agents are related by such a rule.

9. The main theorem

Here's the idea. Suppose that → is a probability conditional for a particular non-trivial probability function. Then we can easily find another probability function for which it is not—indeed, we can exhibit a specific pair of propositions *A* and *B* for which PCCP fails for this function. We use a bold and moderate rule

to shift all probability to a particular proposition that depends on A, B, and $A \rightarrow$ B. Once shifted, the two sides of PCCP are cleaved apart—one side zero, the other positive. The algebraic proof is a bit fiddly because we have to ensure that the relevant conditional probabilities are always defined (though the proofs of Lewis's result and the Import-Export Lemma are even more fiddly in this regard). But the underlying idea is simple.

Main theorem:

Let P be a non-trivial probability function, let $\rightarrow$ be a probability conditional for P, and let $\mathscr{R}$ be a bold[7] and moderate revision rule. Then there exists a proposition X such that $\rightarrow$ is *not* a probability conditional for the function that results from P by accommodating X using $\mathscr{R}$.

Proof:

Let P be non-trivial. Then there are two propositions A and B such that $P(A) < 1$, $P(A \ \& \ B) > 0$ and $P(A \ \& \ \neg B) > 0$. Let $\mathscr{R}$ be a bold and moderate rule. There are two cases:

---

[7] I don't actually need the full strength of the assumption of boldness in the proof. All I need is that, for any non-trivial probability function P, and propositions A and B as described in the proof, the rule can revise P so as to give probability 1 to $(A \ \& \ B) - (A \rightarrow B)$, or so as to give probability one to $\neg(A \ \& \ B)$. That is certainly weaker than boldness, but I don't know a neat way of characterizing it.

Case 1. $(A \& B) - (A \rightarrow B)$ is non-empty[8]
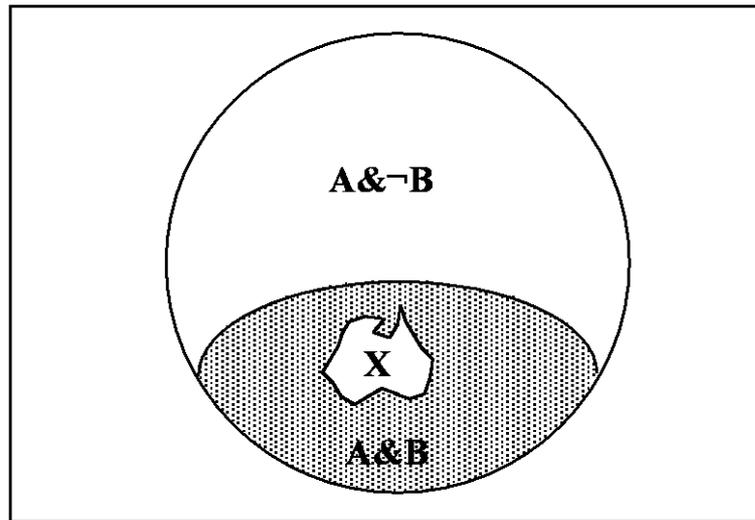
Let $X = (A \& B) - (A \rightarrow B)$. Use $\mathscr{R}$ to revise P to accommodate X, (as we know we can, by $\mathscr{R}$'s boldness) thus producing P_X. Clearly $P_X(A \rightarrow B) = 0$, since we have accommodated X, and X is incompatible with $A \rightarrow B$. But $P_X(B \mid A) = 1$, since $P_X(X) = 1$, and X implies $A \& B$, so that $P_X(A \& B) = 1$, and thus $P_X(A) = 1$, ensuring that this conditional probability is defined. So $\rightarrow$ is not a probability conditional for P_X.

---

[8] As I hope is obvious, $P - Q$ is shorthand for $P \& \neg Q$.

Case 2. $(A \& B) - (A \rightarrow B)$ is empty
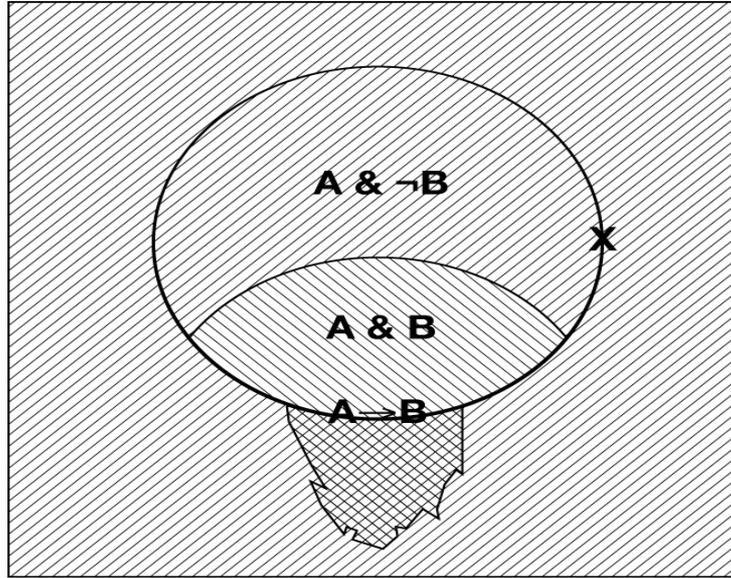
$\neg(A \& B)$ is hatched with lines oriented from bottom left to top right, $A \rightarrow B$ is hatched with lines oriented from top left to bottom right; they intersect in the double-hatched blob that hangs below $A \& B$. (The blob may also invade $A \& \neg B$, although in that case $\rightarrow$ violates modus ponens. I want to allow this since I don't want this result to rest on any assumptions about the logic of $\rightarrow$.)

The blob has positive probability:

$$P([A \rightarrow B] - (A \& B)) > 0 \qquad (1)$$

for otherwise $P(A \rightarrow B) = P(A \& B)$, but then by PCCP, $P(B \mid A) = P(A \& B)$, contradicting our assumption that $0 < P(A \& B)$ and $P(A) < 1$.

Let $X = \neg(A \& B)$. Use $\mathcal{R}$ to revise P to accommodate X (as we know we can, by $\mathcal{R}$'s boldness), thus producing P_X. (1) implies that

$$P_X([A \rightarrow B] - (A \& B)) > 0 \qquad (2)$$

since $[A \rightarrow B] - (A \& B)$ entails X, and $\mathcal{R}$ is moderate. The blob retains positive probability after the revision.

(2) implies that

$$P_X(A \rightarrow B) > 0 \qquad (3)$$

since probability is non-decreasing through entailment. (The blob is part of $A \rightarrow B$.)

Now,

$$P_X(A \ \& \ B) = 0, \qquad\qquad (4)$$

since $P_X(X) = P_X(\neg(A \ \& \ B)) = 1$, while

$$P_X(A) > 0 \qquad\qquad (5)$$

since $P_X(A \ \& \ \neg B) > 0$, by $\mathscr{R}$ being moderate and $A \ \& \ \neg B$ entailing X, and since probability is non-decreasing through entailment. By (4) and (5),

$$P_X(B \mid A) = 0 \qquad\qquad (6)$$

((5) was needed to ensure that this conditional probability is defined.) (3) and (6) imply that $\rightarrow$ is not a probability conditional for P_X.

QED

You may have wondered why I didn't use much the same trick in case 2 as I did in case 1: simply use $\mathscr{R}$ to move all probability onto $(A \rightarrow B) - (A \ \& \ B)$ (instead of onto $\neg(A \ \& \ B)$), so that the probability of $A \rightarrow B$ becomes 1 while the conditional probability of B, given A, becomes 0. In fact, couldn't I have appealed simply to $\mathscr{R}$'s *boldness*, dispensing with its moderateness and strengthening my result accordingly? No—for we do not want the conditional probability to go undefined, as it would if the probability of A became 0. We have good reason to think that $(A \rightarrow B) - (A \ \& \ B)$ is not compatible with A, for their compatibility would require a failure of modus ponens. At a couple of points I needed $\mathscr{R}$ to be moderate to ensure that the relevant conditional probabilities are defined.

Corollary:

If a class of probability functions is closed under a bold and moderate rule, then there is no probability conditional for that class unless the class consists entirely of trivial probability functions.

Proof:

Suppose for reductio that there is a probability conditional $\rightarrow$ for a class $\mathscr{C}$ consisting of at least one non-trivial probability function, and that the class is

closed under a bold and moderate rule $\mathscr{R}$. Choose a non-trivial probability function P from $\mathscr{C}$, and an A and B as in the proof of the main theorem. We may use $\mathscr{R}$ to revise P in order to accommodate either $X = (A \ \& \ B) - (A \to B)$ (case (i)), or $X = \neg(A \ \& \ B)$ (case (ii)), yielding a probability function P_X for which $\to$ is *not* a probability conditional. But by $\mathscr{C}$'s closure under $\mathscr{R}$, P_X belongs to $\mathscr{C}$. This contradicts our assumption that $\to$ is a probability condition for the (entire) class $\mathscr{C}$.   QED

Lewis's second triviality result is a corollary of this corollary, since conditioning is a bold and moderate revision rule. But this corollary is stronger, since there are other such rules. The main theorem is stronger still. And it dispenses with Lewis's somewhat complicated notion of a class of probability functions being closed under a revision rule. Just one (non-trivial) probability function and one application of the rule to that function suffices to produce a violation of PCCP—and we can identify a specific $A \to B$ for which the violation occurs.

10. For aficionados: further generalizations

Lewis (1986) went on to prove a third and fourth triviality result, extending his earlier results to cover different closure assumptions: respectively, closure under conditioning on the members of a single finite partition, and closure under (2-celled) Jeffrey conditioning. As Bas van Fraassen has pointed out to me, we could modify the proof of the main theorem to get a result that has Lewis's third triviality result as a corollary, provided we could assume that contained within (A & B) – ($A \to B$) in Case 1, and within $\neg$(A & B) in Case 2, there is some member of the finite partition, upon which we can conditionalize.

And Aidan Lyon has suggested to me that we could also modify the proof to get a result that has Lewis's fourth triviality result as a corollary. So far we have only considered revision rules that revise probability functions to *fully*

accommodate particular propositions. But we could also consider rules that revise probability functions to *partially* accommodate these propositions, yielding functions that assign them specified probabilities that fall short of 1. Jeffrey conditioning is such a rule, but we could countenance various others. Instead of using a bold revision rule to move *all* probability to designated propositions (which I labeled '*X*'), as I did in the original proof, we could use a *confident* but not-quite-bold rule to move sufficiently large amounts of it, so as still to cleave apart the two sides of PCCP. In the version of the new proof that I have come up with—following Lyon's leads—the moderation of the revision rule needs to be strengthened to an analogue of conservativeness (appropriate for partial accommodation). But the broad outline of the proof is as before, with Case 1 and Case 2 and the designated propositions as in the original proof.

I have decided not to reproduce the new proof here, as it is somewhat tedious, and I have a sense that it is overkill at this stage; this paper is long enough as it stands! I think that the main theorem is already telling against Stalnaker's Thesis. The next section explains why.

11. Vale Stalnaker's Thesis – and weaker theses?

The main theorem shows how *precarious* PCCP is: while it may hold for a single probability function (for all that the theorem tells us), it is easily torn asunder. If you are a Bayesian agent who seeks to conform to PCCP at all times, you are apparently unable to revise boldly and moderately your opinions regarding certain propositions. These propositions then have a curious status for you: you give them positive credence, but you can never learn them—where learning is modeled by a bold and moderate rule. Borrowing terminology from van Fraassen (1984), they are 'Moore propositions'—propositions that you cannot learn without violating a structural constraint that is imposed on you (in this case, the upholding of PCCP). There is something Moore paradoxical about your saying, or thinking: "*p* has positive probability, but it is impossible for me to

learn it", where p may be easily accessible to your enquiry. After all, the propositions—labeled 'X'—that are accommodated in the proof of the main theorem are quite generic.

Similarly, the Bayesian agent cannot *suppose* them, where supposing is modeled by a bold and moderate rule. Collins (1991) and Joyce (1999) argue that supposing should be modeled by imaging; Gärdenfors (1982) generalizes it to blurred imaging. Jaynes (2003) advocates maxent as a revision rule. And Jaeger (1995) favors minxent as a revision rule.

Here, a defender of some version of the Thesis might appeal to *interpretivism* about credences, an influential position in the philosophy of mind: your mental state is the set of probability and utility functions that rationalize your behavioral dispositions as well as possible. We may think of this state as the one attributed to you by an ideal interpreter, whose task is to represent you as charitably as possible. (See Lewis 1974.) If rationality constrains your credences to obey PCCP, then perhaps an ideal interpreter will always attribute to you PCCP-obeying credence functions, in which case your credences *do* in fact always obey PCCP? The trouble is that rationality then apparently may *over*constrain you, for at the same time it may require you to accommodate boldly and moderately a proposition of the kind I designated 'X'. The ideal interpreter is then unable to represent you as entirely rational—either you violate PCCP, or you are diachronically irrational. Lewis already made this ideal interpreter's task an unenviable one with his result about conditioning. To the extent that there other rational ways of revising one's credences boldly and moderately on 'X' propositions, I have only made the interpreter's job harder.

In any case, echoing a remark I made earlier about Lewis's results, the probability functions related by bold and moderate revisions need not be credences of a single agent at different times; they may be those of two different agents at a single time. The upholder of Stalnaker's Thesis, or even various

weaker theses, must insist that at least one of the agents is irrational. This sounds like big news for our theory of rationality.

Since the main theorem extends Lewis's result to a number of other important revision rules, objections merely to conditioning do nothing to undercut my results—the objections would need to carry over to the other rules as well. One objection does: recall van Fraassen's insistence that rationality does not require *rule*-following at all, so by his lights it does not require imaging, or blurred imaging, or revision by maxent, minxent, or any other bold and moderate rule. But while I found it convenient to state the theorem in terms of a revision *rule*, inspection of the proof shows that what matters is the *revision*, not the rule. Wholesale discarding of one's priors and epistemic leaps of faith are just as problematic for the upholding of PCCP as long as they have the effect of shifting all probability onto one of the propositions I designated '*X*'. They may take place even though they are not rule-governed. Surely nothing in van Fraassen's permissive epistemology prevents this. On the contrary, its very permissiveness only makes such shifts all the more permissible!

My result also goes some way to mitigating van Fraassen's concern about the fixed interpretation of the conditional throughout changes in probability functions. It does not go all the way—to be sure, some changes in the identity of a given conditional may sustain PCCP through the sorts of probability revisions that I have envisaged. However, some do not; in fact, some may make matters worse, driving the two sides of PCCP even further apart. For example, where $P_X(A \rightarrow B)$ was positive in my proof, it could be even greater if $A \rightarrow B$ grows through the probability revision, expanding to include all the worlds it previously included and more besides. This would drive $P_X(A \rightarrow B)$ even further away from $0 = P_X(B \mid A)$.

Like Lewis's result, mine assumes nothing about the logic of the conditional beyond its being a 2-place connective. You may replace the '$\rightarrow$' throughout with some less evocative symbol throughout—say, '$\curlyvee$'. There is no

binary operator Υ, conditional-like or otherwise, such that the equation $P(A \Upsilon B)$ = $P(B \mid A)$ survives all revisions by a given bold and moderate revision rule. Conditional probability is not a special kind of unconditional probability, however one might construe that kind. It's the probability that drives my result rather than the logic.

But in fact my result needs virtually no probability theory either. For example, it does not assume that probabilities are additive, nor the law of total probability. Of course, to the extent that the Thesis is one concerning the *probabilities* of conditionals, we should not begrudge these assumptions. Still, it is all the better if we can do without them, as I have.

12. Transmogrifying Stalnaker's Thesis into Adams' Thesis

The last observation brings to mind Adams' Thesis, the variant of Stalnaker's Thesis, applied to indicative conditionals that lack truth values. Lewis (1976) concedes that Adams need not be troubled by his triviality results:

> 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 applying standard laws of probability to the probabilities of conditionals…
>
> 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. (303 - 305)

It might look like my result can claim another victim, scotching any interesting version of Adams' Thesis. With my assuming so little about 'P', perhaps it can

just as well be interpreted as representing "assertability", in which case it is still answerable to the main theorem?

Alas, no. For recall that the reason that Adams-style assertabilities do not obey the probability calculus is that their domain is not a field. In particular, Adams assumes that indicative conditionals cannot be negated or conjoined with other sentences. So Adams would baulk at Boolean combinations such as $(A \ \& \ B) - (A \rightarrow B)$ or $(A \rightarrow B) - (A \ \& \ B)$, the centrepieces of my proof.

In my opinion, far from saving Adams's position from trivialization, this only reveals its incompleteness. For English takes such combinations in its stride. "If Oswald didn't shoot Kennedy then somebody else did, but it's not the case that both those things happened—so Oswald must have shot Kennedy" seems perfectly felicitous. More generally, Adams must face the familiar 'Frege-Geach'-style objection that indicative conditionals compound truth-functionally with other sentences. That said, various 'no-truth-value' theorists in the Adams tradition, such as McGee (1989), Edgington (1995) and Bennett (2003) face up to this problem, and this is still an area of lively debate. Indeed, the very triviality results against Stalnaker's Thesis motivate their no-truth-value accounts. The upshot is that my result here does not scathe Adams' Thesis.

I have another triviality result that I believe does. It is based on my (1989) theorem, strengthened in Hájek and Hall (1994), that PCCP cannot hold for any finite-ranged probability function. I show that for any such function, there are more distinct conditional probability values than there are distinct unconditional probability values, so there is no way of matching all the former with all the latter. *A fortiori*, there is no way of matching all the distinct conditional probability values with probability-of-conditional values—there will always be some unmatched conditional probability values. If Adams' Thesis about assertability is to hold, then assertability must be more nuanced than unconditional probability: conditional probability values outrun unconditional probability values, so if assertability values are to keep pace, they too must outrun

unconditional probability values. I doubt that this is the case: if anything, the latter outrun the former. But this is just a promo for a coming attraction; for the full story, see this paper's sequel (Hájek forthcoming).

13. Final scene

For now, I am content to have chased down and cornered Stalnaker's Thesis some more. I have closed off more of its escape routes. It appears to be moribund. But as aficionados of the *Friday the 13*[th] series know, you should never be surprised if Jason suddenly reappears in a new guise—and the same is true of Stalnaker's Thesis.

Research School of Social Sciences
Australian National University
Canberra, ACT 0200, Australia
alan.hajek@anu.edu.au

REFERENCES

Adams, Ernest (1975): *The Logic of Conditionals*, Dordrecht: Reidel.

Bacchus, F, H. E. Kyburg, and M. Thalos (1990): 'Against Conditionalization', *Synthese* 85, 475-506.

Bennett, Jonathan (2003): *Conditionals*, Oxford: Oxford University Press.

Cartwright, Nancy (1999): *The Dappled World: A Study of the Boundaries of Science,* Cambridge: Cambridge University Press.

Collins, John (1991): *Belief Revision*, Ph.D. dissertation, Princeton University.

de Finetti, Bruno (1972): *Probability, Induction and Statistics*, Wiley.

Edgington, Dorothy (1995): "On Conditionals", *Mind* 104 (414), 235-329.

Eells, Ellery, and Brian Skyrms (eds.) (1994): *Probability and Conditionals*, Cambridge: Cambridge University Press.

Gärdenfors, Peter (1982): "Imaging and Conditionalization", *Journal of Philosophy* 79, 747-760.

Gibbard, Alan and William Harper (1978): "Counterfactuals and Two Kinds of Expected Utility", in C. A. Hooker, J. J. Leach, and E. F. McClennen, eds., *Foundations and Applications of Decision Theory*, Vol. 1, Dordrecht: Reidel.

Hájek, Alan (1989): "Probabilities of Conditionals—Revisited", *Journal of Philosophical Logic* 18, 423-428.

Hájek, Alan (1994): "Triviality on the Cheap?", in Eells and Skyrms (1994).

Hájek, Alan (forthcoming): "The Fall of Adams' Thesis?", *Journal of Language, Logic and Information*.

Hájek, Alan and Ned Hall (1994): "The Hypothesis of the Conditional Construal of Conditional Probability", in Eells and Skyrms (1994).

Hall, Ned (1994): "Back in the (CCCP)", in Eells and Skyrms (1994).

Harper, W. L., Robert Stalnaker and Glenn Pearce. (eds.) (1981): *Ifs*, Dordrecht: Reidel.

Harper, William L., and Brian Skyrms (1988): Introduction to *Causation, Chance, and Credence*, Dordrecht: Kluwer.

Hild, Matthias (1998): "The Coherence Argument Against Conditionalization", *Synthese* 115, 229-258.

Jaeger, Manfred (1995): "Minimum Cross Entropy Reasoning" International Joint Conference On Artificial Intelligence, *Proceedings of the 14th international joint conference on Artificial intelligence* - Volume 2, 1847-1852.

Jaynes, Edwin Thompson (2003): *Probability Theory: The Logic of Science*, Cambridge: Cambridge University Press.

Jeffrey, Richard (1966): *The Logic of Decision*, Chicago: University of Chicago Press; 2nd ed. 1983.

Joyce, James M. (1999): *The Foundations of Causal Decision Theory*. Cambridge: Cambridge University Press.

Levi, Isaac (1980): *The Enterprise of Knowledge,* MIT Press, Cambridge, Massachussetts.

Lewis, David (1973): *Counterfactuals*, Blackwell and Harvard University Press.

Lewis, David (1974): "Radical Interpretation", *Synthese* 27: 331-44; reprinted in *Philosophical Papers*, Vol 1, New York and Oxford: Oxford University Press 1983.

Lewis, David (1976): "Probabilities of Conditionals and Conditional Probabilities", *Philosophical Review* 85, 297-315; reprinted in Harper et al.

Lewis, David (1981): "Causal Decision Theory", *Australasian Journal of Philosophy* 59, 5-30; reprinted in Lewis (1986a).

Lewis, David (1986): "Probabilities of Conditionals and Conditional Probabilities II", *Philosophical Review* 95, 581-589.

Lewis, David (1986a): *Philosophical Papers*, Volume II, Oxford University Press.

McGee, Vann (1989): "Conditional Probabilities and Compounds of Conditionals", *Philosophical Review* 98, No. 4, 485-541.

Milne, Peter (2003): "The Simplest Lewis-Style Triviality Proof Yet?", *Analysis* 63.4 (October), 300-303.

Ramsey, F. P. (1931/1990): "Truth and Probability", in *Foundations of Mathematics and other Essays*, R. B. Braithwaite (ed.), London: Routledge

& P. Kegan, 1931; reprinted in *Philosophical Papers*, D. H. Mellor (ed.) Cambridge: Cambridge University Press, 1990.

Ramsey, Frank Plumpton (1965): *The Foundations of Mathematics (and Other Logical Essays)*, Routledge and Kegan Paul.

Rehder, Wulf (1982): "Conditions for Probabilities of Conditionals to be Conditional Probabilities, *Synthese* 53, 439-443.

Savage, Leonard J. (1954): *The Foundations of Statistics,* New York: John Wiley.

Skyrms, Brian (1980): *Causal Necessity*, New Haven: Yale University Press.

Skyrms, Brian (1984): *Pragmatics and Empiricism*, New Haven: Yale University Press.

Sobel, Jordan Howard (1994): *Taking Chances: Essays on Rational Choice*. New York: Cambridge University Press.

Stalnaker, Robert (1968): "A Theory of Conditionals", *Studies in Logical Theory*, American Philosophical Quarterly Monograph Series, No. 2, Oxford: Blackwell.

Stalnaker, Robert (1970): "Probability and Conditionals", *Philosophy of Science* 37, 64-80; reprinted in Harper et al.

Stalnaker, Robert (1976): "Letter to van Fraassen", in Harper and Hooker (eds.), *Foundations of Probability Theory, Statistical Inference and Statistical Theories of Science*, Vol. I, Dordrecht: Reidel, 302-306.

Titelbaum, Michael G. (2008): "The Relevance of Self-Locating Beliefs", *Philosophical Review* 117, 555-606.

van Fraassen, Bas (1976): "Probabilities of Conditionals", in Harper and Hooker (eds.), *Foundations of Probability Theory, Statistical Inference and Statistical Theories of Science*, Vol. I, Dordrecht: Reidel, 261-301.

van Fraassen, Bas (1984): "Belief and the Will", *Journal of Philosophy* 81, 235-256.

van Fraassen, Bas (1989): *Laws and Symmetry*, Oxford: Clarendon Press.