General Dynamic Triviality Theorems

Jeffrey Sanford Russell and John Hawthorne

December 2015 draft

These principles have a ring of plausibility:

> The probability you assign the conditional "If A then B" ought to be the same as the conditional probability you assign B given A.
>
> You ought to believe that A is good to the same degree that you desire that A be the case.
>
> Conditional on "Probably A", you ought to be more confident in A than not-A.
>
> Conditional on it being better for A to be the case than B, you ought to desire A to a higher degree than B.

David Lewis proved famous "triviality theorems" about the first two of these principles (1976; 1986; 1988; 1996). In the standard framework of probability and decision theory, each of them has unacceptable trivializing consequences.

We have three goals in this paper. In Sections 1–3, we generalize Lewis's results to apply to a wide variety of other principles that have a similar flavor—principles about epistemic "might", probability claims, and claims about comparative value, including the last two in the list above. In Sections 4–5, we show how to streamline these arguments in a more minimalistic framework, which doesn't depend on standard conditionalization or numerical probabilities. Instead, the arguments turn on structural principles about updating with new information—namely, that updates are *commutative* in the sense that the order of information doesn't matter, and *idempotent* in the sense that repeatedly updating on the same information is redundant.[1] In Section 7, we explore the prospects for maintaining some or all of the problematic principles by giving up these structural principles about update, drawing

Thanks to Andrew Bacon, Sarah Moss, and several anonymous referees for helpful comments.

[1] These streamlined results are connected to existing "qualitative" triviality results in belief-revision frameworks such as Gärdenfors (1986). But there are important differences between their assumptions and ours, which we'll discuss when we get there.

on the resources of dynamic semantics. Many philosophers (including Lewis, following Ernest Adams 1975) have been attracted to an "expressivist" reply to the triviality results for conditionals, rejecting truth conditions for conditionals. Analogous positions are also natural for epistemic modals or evaluatives. We argue that the dynamic approach is a more promising version of this expressivist idea than the version Adams and Lewis influentially advocated. Along the way (Section 6) we also consider a "contextualist" proposal for maintaining parameterized versions of these principles, and raise a general difficulty for it arising from anti-luminosity considerations.

(The more abstract arguments in Sections 4 and 5 don't strictly depend on the earlier material, so while the gentler course is to read straight through, some readers might prefer to take an accelerated track starting there.)

1 Probabilities of Conditionals

Lewis's (1976) triviality theorem[2] addresses a principle about what subjective probability—or *credence*—it is reasonable to assign a conditional:

Stalnaker's Thesis. For every reasonable credal state C, if $C(A) > 0$,

$$C(\text{if } A, B) = C(B \mid A)$$

Here $C(B \mid A)$ is the conditional credence of B given A. This is standardly defined as the ratio $C(A \wedge B)/C(A)$. The qualification $C(A) > 0$ is necessary because the ratio is undefined when $C(A) = 0$. (Note that for the next few sections we'll be following Lewis in several technically standard probabilistic assumptions—including the standard ratio definition of conditional probability, which doesn't make sense for evidence with prior credence zero. But later we'll examine how things go when these assumptions are relaxed. It will turn out that not very much turns on them.)

We suppose that if C is reasonable and $C(A) > 0$, then the credal state that results from C by conditionalizing on A is also reasonable. (This is the function C_A such that $C_A(B) = C(B \mid A)$ for each proposition B.)

[2]There is of course a whole cottage industry of other triviality results for conditionals, using different assumptions from Lewis's. (Some of the main ones: Stalnaker 1976a; Hájek and Hall 1994; Edgington 1986; Hájek 1989.) We won't be engaging very much with these other results. One reason for this is that we'll be looking at generalizations to other domains on which these other results don't bear so obviously. Another reason is that it turns out that the "core" version of Lewis's original triviality result is already more powerful than has generally been appreciated.

We'll say that propositions A and B are **compatible** (for credences C) iff $C(A) > 0$ and $C(B \mid A) > 0$; otherwise they are **incompatible**. Here are two important standard probabilistic facts about this notion of compatibility.

> **Negation.** For any reasonable credences, $\neg A$ and A are incompatible.
>
> **Symmetry.** For any reasonable credences, if A and B are compatible, then B and A are compatible.

(This is easy to check using the ratio definition.) Note that given Symmetry and the definition of compatibility, if follows that if $C(B) = 0$ then A and B are incompatible. We'll also say A **implies** B iff A is incompatible with $\neg B$.

Lewis derives from Stalnaker's Thesis the surprising conclusion that contingent conditionals imply their consequents: that is, for any propositions A and B, either A is incompatible with $\neg B$, or else (if A, B) implies B. He goes on to draw other absurd consequences from this fact. (For instance: no three propositions are possible but pairwise incompatible.) But note that this consequence is already bad enough. For example, you might not be sure whether Andrew is at the talk, nor whether if Andrew is at the talk he's happy. But when you learn the conditional—if Andrew is at the talk, he's happy—this still generally leaves you in the dark about whether Andrew is in fact happy. So Lewis's extra bells and whistles aren't necessary for making trouble for Stalnaker's Thesis.

Here's a simplified version of the argument. Consider any reasonable credal state C for which A and $\neg B$ are compatible, and thus by Symmetry, $\neg B$ and A are compatible. Let $C_{\neg B}$ be the credal state that results from conditionalizing C on $\neg B$; this is another reasonable state, and

$$C_{\neg B}(A) = C(A \mid \neg B) > 0$$
$$C_{\neg B}(B) = C(B \mid \neg B) = 0$$

(The first is because $\neg B$ is compatible with A for C, and the second follows from Negation.) Since $C_{\neg B}(B) = 0$, it follows that A is incompatible with B for $C_{\neg B}$, and so $C_{\neg B}(B \mid A) = 0$. Then Stalnaker's Thesis implies

$$C(\text{if } A, B \mid \neg B) = C_{\neg B}(\text{if } A, B) = C_{\neg B}(B \mid A) = 0$$

That is, $\neg B$ is incompatible with (if A, B) for C. Then by Symmetry, (if A, B) is incompatible with $\neg B$, which is to say that the conditional implies its consequent. QED.

2 Desire as Belief

Lewis's second "triviality" result is about subjective values. Let's start by saying a bit about what these are like.

Somebody flipped a coin, and if it came up heads they put \$100 in Box A; if it came up tails, they put in nothing. Box B contains \$80. One of the boxes is yours. Which box would you rather it be? There's one sense in which, if unbeknownst to you the coin came up heads, Box A being yours is better than Box B being yours. But for you in your ignorance, there's a "subjective" sense in which it's better news to find out you got Box B.

Lewis follows Richard Jeffrey (1983 [1965]) in representing subjective values with a function V from propositions to degrees of desirability. For a proposition A, the value $V(A)$ is how desirable you expect reality to be, conditional on A. For a tautology $\top$, $V(\top)$ is just how desirable you expect reality to be given your current information. So in general $V(A)$ is equal to $V_A(\top)$, where (C_A, V_A) is the result of conditionalizing (C, V) on A.[3] In the case where $C(A) = 0$, conditionalization on A is standardly undefined, and so $V(A)$ is undefined as well.

The target of Lewis's triviality argument is the "anti-Humean" view that there are necessary connections between subjective value and beliefs about what is good.

> Then the Desire-as-Belief thesis says that Frederic desires things just when he believes they would be good. Or better, since we must acknowledge that desire and belief admit of degree, he desires things just to the extent that he believes they would be good. (1988, 325–26)

So for a proposition A we have a corresponding proposition $(\mathrm{good}\, A)$, such that the following correspondence holds:

> **Desire as Belief.** For every reasonable combination of a credal state C with an evaluative state V, if $C(A) > 0$ then
>
> $$C(\mathrm{good}\, A) = V(A)$$

As Lewis and others have shown, this thesis is disastrous. We'll present a simplified argument, which deploys the following key fact about standard conditionalization:[4]

[3] The definition of conditionalization for a credal-evaluative pair (C, V) and a proposition A is the pair (C_A, V_A) such that, for each proposition B, (i) $C_A(B) = C(A \wedge B)/C(A)$, and (ii) $V_A(B) = V(A \wedge B)$.

[4] Lewis's 1988 result was more complicated, and involved assumptions about Jeffrey conditionalization. Our presentation is closer to the simpler arguments given by Arló Costa, Collins, and Levi (1995) and Lewis (1996), though it is not quite the same as either of those.

Commutativity. Let (C, V) be a reasonable credal-evaluative pair, and let A and B be propositions which are compatible for C. Let (C_A, V_A) be the result of conditionalizing (C, V) on A, and let (C_B, V_B) be the result of conditionalizing (C, V) on B. Then the result of conditionalizing (C_A, V_A) on B is the same as the result of conditionalizing (C_B, V_B) on A. More concisely:

$$(C_A)_B = (C_B)_A$$
$$(V_A)_B = (V_B)_A$$

The Commutativity principle thus stated is a theorem of standard decision theory. In contrast, other epistemological claims that go by the same name are contested. For a representative example, Frank Arntzenius offers this thesis: "if the relevant evidence and experience collected is the same, then the order of collection should not matter for the final degrees of belief" (2003, 364). Of course, on the supposition that "the final degrees of belief" are always calcluated by standard conditionalization, Arntzenius's thesis would follow from the theorem. But that supposition is highly controversial. We will return to these issues in Section 4.

Like the argument against Stalnaker's Thesis, this argument also relies on a more substantive, but very natural assumption about conditionalization: if (C, V) is a reasonable pair of states for which $C(A) > 0$, then (C_A, V_A) is also reasonable.

Given these assumptions, we can derive a surprising Pessimistic Conclusion: for any reasonable state (C, V) for which A and $(\text{good}\, A)$ each have non-zero credence, $\neg\, \text{good}\, A$ implies A. In other words, whatever is not good, must be the case. (Lewis draws other trivializing consequences from the Desire as Belief thesis. But the Pessimistic Conclusion looks bad enough.)

Here is the argument. Let (C, V) be a reasonable credal-evaluative pair for which $C(A)$ is non-zero, and let B be any proposition. Note that by the probability calculus,

$$C_A(A \vee B) = C(A \vee B \mid A) = 1$$

Moreover, conditionalizing on what you already are certain of doesn't change your attitude. So $(V_A)_{A \vee B} = V_A$, and thus by Commutativity, $(V_{A \vee B})_A = V_A$. Thus:

$$V_{A \vee B}(A) = (V_{A \vee B})_A(\top) = V_A(\top) = V(A)$$

By two applications of Desire as Belief:

$$\begin{aligned} C(\text{good}\, A \mid A \vee B) = C_{A \vee B}(\text{good}\, A) &= V_{A \vee B}(A) \\ &= V(A) = C(\text{good}\, A) \end{aligned}$$

This much shows that Desire as Belief requires that $\mathsf{good}\,A$ is probabilistically *independent* of $A \vee B$, for any proposition B. But suppose that B is itself the proposition $(\mathsf{good}\,A)$: then B is independent of $A \vee B$, and this can only hold if $\neg B$ implies A. (This follows because independence is *symmetric*, and so $C(A \vee B) = C(A \vee B \mid A) = 1$, and thus $C(\neg A \wedge \neg B) = 0$.) That is, $\neg\,\mathsf{good}\,A$ implies A. QED.

3 Triviality Theorems Galore

Now that we've got the hang of these core arguments, we can extend the same ideas to make trouble for other plausible-sounding theses. In this section we'll continue to assume that if (C, V) is a reasonable pair of credal and evaluative states, then so is (C_A, V_A), if $C(A) > 0$.

Let's start with "probably". Here's a thesis with a similar kind of appeal to Stalnaker's Thesis: conditional on the claim that A is probably the case, you should assign A higher probability than $\neg A$.

> **Probably.** For any reasonable credal state C, if $C(\mathsf{probably}\,A) > 0$, then
>
> $$C(A \mid \mathsf{probably}\,A) > {}^1\!/_2$$

But Probably has the trivializing consequence that $(\mathsf{probably}\,A)$ implies A. As in Lewis's proofs, the trick for showing this is to update on the right piece of evidence: in this case, $\neg A$. Suppose for reductio that for some reasonable credal state C, $\neg A$ and $\mathsf{probably}\,A$ are compatible. This tells us that if $C_{\neg A}$ is the result of conditionalizing C on $\neg A$, then

$$C_{\neg A}(\mathsf{probably}\,A) = C(\mathsf{probably}\,A \mid \neg A) > 0$$

Applying Probably to $C_{\neg A}$, we can then conclude

$$C_{\neg A}(A \mid \mathsf{probably}\,A) > {}^1\!/_2$$

But also,

$$C_{\neg A}(A) = C(A \mid \neg A) = 0$$

This implies that $\mathsf{probably}\,A$ is incompatible with A for $C_{\neg A}$, and so

$$C_{\neg A}(A \mid \mathsf{probably}\,A) = 0 < {}^1\!/_2$$

This is a contradiction. So it must be that our supposition was false, and $\neg A$ and $(\mathsf{probably}\,A)$ are incompatible for C. Then by Symmetry, $(\mathsf{probably}\,A)$ is incompatible with $\neg A$, which is to say that $(\mathsf{probably}\,A)$ implies A.

Exactly the same style of proof makes trouble for other principles about probability claims. For instance, this also has the ring of plausibility: conditional on the claim that the probability of A is x, you should assign credence x to A.[5]

Exact Probability. For any reasonable credal state C, if

$$C(\text{the probability of } A \text{ is } x) > 0$$

then

$$C(A \mid \text{the probability of } A \text{ is } x) = x$$

But this implies that for any $x > 0$, (the probability of A is x) implies A, by updating on $\neg A$ and reasoning in the very same way.

Here's another. If you learn that A is more probable than B, you should assign higher credence to A than B.

Comparative Probability. For any reasonable credal state C, if

$$C(A \text{ is more probable than } B) > 0$$

then

$$C(A \mid A \text{ is more probable than } B) > C(B \mid A \text{ is more probable than } B)$$

But again, conditionalizing on $\neg A$ shows us that (A is more probable than B) implies A. (In this case we could also draw the conclusion that (A is more probable than B) is incompatible with B, using a similar argument involving conditionalizing on B.)

Using the same technique, we can make trouble for similarly attractive principles concerning epistemic modals that aren't explicitly probabilistic. The simplest case is epistemic "might". Here's the target principle:

Might. For any reasonable credal state C, if $C(\mathsf{might}\,A) > 0$, then

$$C(A \mid \mathsf{might}\,A) > 0$$

The same style of proof shows that if Might is true, then $(\mathsf{might}\,A)$ implies A.[6]

[5] Note this principle has the same structure as a (synchronic) reflection principle. (For discussion see e.g. Christensen 2007; 2010.)

[6] Yalcin (2007) discusses a related puzzle about epistemic "might": it is difficult to understand the supposition of "not-A and it might be that A", but also, not-A and $\mathsf{might}\,A$ are not straightforwardly incompatible. His response broadly falls along the lines of the positions we consider in Sections 4 and 7.

Similar considerations apply to many plausible-sounding theses about subjective value. We'll frame these theses in a way that avoids a couple of oddities of Lewis's formulation of Desire as Belief, the equation $C(\text{good} A) = V(A)$. First, note that it's a bit odd to put credences and values on the same numerical scale. After all, credences are bounded by zero and one, but values are plausibly unbounded, and there's no obvious sense to a zero-point or a unit on the scale of desirability.[7] Furthermore, the Desire as Belief thesis conflates two intuitively separate concerns: the probability of something being good, and how good it might be. It could be that getting a box of unknown contents is very likely to just pass the threshold of goodness, but this doesn't require it to have a high degree of desirability. It may be certain to at best *just* pass the threshold for goodness (say containing a pair of socks). Moreover, it could be that in the unlikely event that getting the box is bad, it is really terrible (say containing a deadly python). Then even though it's probable that getting the box is good, it is subjectively highly undesirable.

These are problems for theses that directly link up credences about values to degrees of desirability. We'll set aside Lewis's particular form of Desire as Belief. But there are other interesting theses linking desire and belief that don't take this form. Consider the proposition that A is more desirable than B. Suppose you learn this. It seems plausible that in this case, you ought to assign higher subjective value to A than you do to B. This suggests the following principle:

> **Comparative Value.** Let (C, V) be any reasonable state such that
>
> $$C(A \text{ is better than } B) > 0$$
>
> and let (C_+, V_+) be the result of updating with (A is better than B). Then
>
> $$V_+(A) > V_+(B) \qquad \text{unless } C_+(A) = 0 \text{ or } C_+(B) = 0$$

Again, this principle has unacceptable consequences. We'll present an argument that relies on Commutativity together with the following further fact about standard conditionalization:

> **Idempotence.** Let (C, V) be a reasonable credal-evaluative pair for which $C(A) > 0$. Then if (C_A, V_A) is the result of conditionalizing on A, the result of conditionalizing (C_A, V_A) on A is the very same pair. More concisely:
>
> $$(C_A)_A = C_A$$
> $$(V_A)_A = V_A$$

[7] One manifestation of this point is the problem utilitarians face of making sense of interpersonal comparisons of desirability; for discussion see e.g. Hausman (1995).

The idea behind this is that if you have conditionalized on A, then you're certain of A, and conditionalizing on what you already certain of does not change your attitudes.

Suppose (for reductio) that (C, V) is a reasonable state for which A, B, and (A is better than B) are compatible. Then $((C_A)_B, (V_A)_B)$ is a reasonable pair, and we can conditionalize this on (A is better than B). As a shorthand, let

$$\begin{aligned} C_{AB+} &= ((C_A)_B)_{A \text{ is better than } B} \\ V_{AB+} &= ((V_A)_B)_{A \text{ is better than } B} \end{aligned}$$

That is, (C_{AB+}, V_{AB+}) is the result of conditionalizing (C, V) on A, then on B, then on (A is better than B). Comparative Value tells us

$$V_{AB+}(A) > V_{AB+}(B)$$

On the other hand,

$$\begin{aligned} (V_{AB+})_A &= (((V_A)_B)_{A \text{ is better than } B})_A \\ &= (((V_A)_A)_B)_{A \text{ is better than } B} && \text{by Commutativity} \\ &= ((V_A)_B)_{A \text{ is better than } B} && \text{by Idempotence} \\ &= V_{AB+} \end{aligned}$$

By similar reasoning, $(V_{AB+})_B = V_{AB+}$. So

$$V_{AB+}(A) = (V_{AB+})_A(\top) = (V_{AB+})_B(\top) = V_{AB+}(B)$$

This contradicts the inequality stated above. So Comparative Value entails that, for every reasonable pair (C, V), A, B, and (A is better than B) are jointly incompatible. But this is a very strange result. For instance, suppose A is the proposition that you'll receive a million dollars, and B is the proposition that you'll have a pleasant lunch. A being better news than B doesn't preclude both being true.

Similar arguments apply to some other plausible-sounding principles about subjective value. Setting aside Lewis's somewhat idiosyncratic version of Desire as Belief, here is a more minimal principle about the proposition that A is good. If you learn this, then A should be good news for you: that is, you should assign higher subjective value to A than to the status quo—your value for A should exceed your value for, say, $2 + 2 = 4$.

> **Good News.** Let (C, V) be a reasonable state for which $C(\mathsf{good}\, A) > 0$, and let (C_+, V_+) be the result of updating with $\mathsf{good}\, A$. Then for a tautology $\top$,
>
> $$V_+(A) > V_+(\top) \qquad \text{unless } C_+(A) = 0$$

For the same reason, there is bad news for Good News. Just substituting $\top$ in for B in the argument about Comparative Value, we conclude that for any reasonable state, $\mathsf{good}\,A$, A, and $\top$ are jointly incompatible. Since $(C_\top, V_\top) = (C, V)$, this means that $\mathsf{good}\,A$ and A are incompatible. So Good News implies Converse Pessimism: what is good cannot be true.

Consider one last principle. Suppose you learn that A is desirable to degree x. In that case, it seems plausible that the subjective value you ought to assign A is x.[8]

> **Exact Value.** Let (C, V) be a reasonable state for which
>
> $$C(\text{the value of } A \text{ is } x) > 0$$
>
> and let (C_+, V_+) be the result of adding as evidence (the value of A is x). Then
>
> $$V_+(A) = x \qquad \text{unless } C_+(A) = 0$$

By a similar line of reasoning, this principle implies that for any distinct values x and y, the four propositions A, B, (the value of A is x), and (the value of B is y), are jointly incompatible. Again, this is a very strange result. The information that there is ten dollars in Box A and twenty dollars in Box B surely doesn't rule out the possibility that you will get both boxes. (We defer the proof of this fact until after we introduce simpler notation in Sections 4 and 5.)

4 No Truth Value?

Let's return to conditionals. A standard reply to Lewis's triviality theorem is to say that Stalnaker's Thesis can be retained if we maintain that conditional sentences "have no truth values, no truth conditions, and no probabilities of truth" (Lewis 1976, 303).[9] Of course, the argument didn't have any premises explicitly about *truth*. So for this idea to help escape triviality, it needs to be elaborated. As Lewis points out:

> Merely to deny that probabilities of conditionals are probabilities of truth, while retaining all the standard laws of probability in suitably adapted form,

[8]Here we could assume some arbitrary numerical scale of value, and let x be a number on that scale. But there's no need to do that: x could just as well be a quantity of desirability itself, rather than a number.

[9]The "expressivist" idea that certain declarative sentences lack truth values is argued for independently in each of the domains we've discussed. On conditionals, for overview see Bennett (2003, sec. 38) and Edgington (2014, sec. 3). On probability operators, see Yalcin (2012). On evaluatives the literature is vast; see van Roojen (2014) for an overview.

> would not yet make it safe to revive the thesis that probabilities of conditionals are conditional probabilities. It was not the connection between truth and probability that led to my triviality results, but only the application of standard probability theory to the probabilities of conditionals.

Accordingly, Lewis goes on to offer two ways someone might modify standard probability theory.

> He might drop the requirement that the domain of a probability function is a Boolean algebra, in order to exclude conjunctions with conditional conjuncts from the language. Or he might instead limit … the law of additivity, refusing to apply it when the disjuncts A and B contain conditional conjuncts. Either maneuver would block my proofs (1976, 304).[10]

But in fact, neither maneuver looks like it blocks our simplified proof. Look again, and you'll see that the proof never conjoined a conditional with anything, and never appealed to the law of additivity.[11]

Of course, we did use certain facts about conditional credences and conditionalization, so it's natural to worry that we might have illicitly smuggled in some such assumptions about conjunctions or additivity, in which case the streamlined argument could be blocked by one of these maneuvers after all. Indeed, the standard definition of conditional probability makes reference to the probability of a conjunction: for instance, the credence $C(\text{if } A, B \mid \neg B)$ is standardly identified with the ratio of credences

$$\frac{C(\neg B \wedge (\text{if } A, B))}{C(\neg B)}$$

This ratio involves a conjunction that Lewis's first maneuver might exclude.

But identifying conditional credences with ratios of unconditional credences is independently problematic. An ideally sharp dart thrown at a square region has zero

[10] These maneuvers have become standard responses to the triviality results. Adams (1975) identifies the assumption to be rejected thus: "*[T]he probability of a proposition is the same as the probability that it is true.* A more exact statement is that the probability of a proposition is the sum of the probabilities of the possible states of affairs in which the proposition would be true …" (p. 2, original emphasis). That is, for Adams giving up probabilities of *truth* amounts to giving up the additive law for probabilities. Proposing restrictions on embedding conditionals is also a common reaction. See Bennett (2003, sec. 39) for overview.

[11] There's a more radical way of developing the claim that conditionals don't have truth conditions: by denying that conditional sentences express anything which is a fit object of credence at all, or which can be learned or supposed. (And of course one might take the same line on epistemic modals or evaluatives.) This approach doesn't provide a way of maintaining Stalnaker's Thesis (or the other theses we've considered). For that thesis says *which* credence you ought to assign a conditional, and thus implies that you ought to assign *some* credence to a conditional. Since we're specifically interested in "expressivism" as a way of defending these theses from triviality, we'll set this radical version aside.

probability of landing on the left edge (E); but even so, it might do so, and conditional on it hitting the edge it seems sensible to assign probability $^1\!/_2$ to it hitting the upper half (H). In this case, the probability of E is zero and the ratio $C(H \wedge E)/C(E)$ is thus undefined—but the conditional probability $C(H \mid E)$ remains perfectly intelligible (Hájek 2003). Furthermore, zeros aside, even if there is no conjunction with a conditional conjunct, and thus the ratio formula for $C(\text{if } A, B \mid E)$ is undefined, it still makes perfect sense to ask how your credence in the conditional should evolve when you get new evidence; thus it still makes sense to assign it a conditional credence given such evidence. So someone who denies the existence of conjunctions with conditional conjuncts has an additional reason to reject the identification of conditional credence with a ratio of unconditional credences.

Since we shouldn't appeal to the standard ratio definition of conditional credence, it's important for us to clarify what structural assumptions are involved in our arguments. There are standard ways of handling the mathematics of primitive conditional probabilities (such as Popper functions); but in fact the assumptions we need are even thinner than those built into such frameworks. For example, we won't assume that an epistemic state involves assigning numerical credences to any propositions. So we'll restate the triviality proof now using a simple, somewhat nonstandard technical framework.

We'll take as primitive the notion of a **reasonable state** S, and the notion of **updating** a state on a certain proposition A; we label the state that results from this update $S[A]$. These states are to be understood as abstract representations of certain epistemically important features of agents (in the same spirit as credence functions, or the "information states" of e.g. Veltman 1996, which we discuss in Section 7). We are neutral for now as to exactly which features are represented, but states shouldn't be understood as fully detailed psychological descriptions. Something that makes a psychological difference between agents may, even so, make no difference to what "state" they count as being in.

This is a good place to approach more carefully a point we've kept in the background so far. We've sometimes glossed conditional probabilities as the probabilities that one ought to have after learning a certain proposition. For many purposes this gloss is unproblematic, but it may not generally be the best way to understand what updating an epistemic state with a proposition really comes to.

One concern arises from standard objections to Ramsey's famous test for conditionals: "If two people are arguing 'If *p* will *q*?' and are both in doubt as to *p*, they are adding *p* hypothetically to their stock of knowledge and arguing on that basis about *q*" (1990 [1929], 155). There are cases where this way of evaluating a conditional seems to give wrong results. You can be confident that if you're being spied on you don't know it. But on the hypothesis that the proposition that you're being spied on was added to your stock of knowledge, you would know it. (This kind of example is attributed to Richmond Thomason; see Bennett 2003, 28–29.) Similarly, you may assign high conditional probability to not knowing you're being spied on, conditional on being spied on. But we might think that after learning that you're spied on, you ought not assign high probability to not knowing it.

It's not so obvious that this line of thought shows that it's wrong to equate the conditional probability of H on E with the probability you ought to have in H upon learning E—and nothing more. Normally when one learns that one is being spied on, one also notices having learned that. That just means that learning you're being spied on and *nothing more* isn't a normal case.

But there are other problems for the Ramsey Test that carry over more straightforwardly. Take the proposition M that it's raining and you never learn it. Conditional on M, it's sensible to have high conditional probability that you won't get to hear today's weather report. But the hypothesis that you learn M is incoherent (given that learning is factive and distributes over conjunction: see Fitch 1963, Theorem 4; see also Chalmers and Hájek 2007).

A better way to think about updating on new evidence is not in terms of learning, but rather in terms of *supposing*. While you can't learn that it's raining and you never learn it, it's easy to suppose this, and to consider what to think about other questions given this supposition. So we'll think of $S[A]$ as the epistemic state that results from adding to the state S the supposition that A. For simplicity, like standard Bayesians we assume that for any state S there is at most one reasonable state that results from adding a particular proposition as evidence.

But there may still be suppositions which do not result in any reasonable state—for example, logical contradictions. Or if you suppose that Bea was born in 1985, then the additional supposition that Bea was born in 1980 leads to absurdity. In general, let's say that S **rules out** A iff $S[A]$ is not a reasonable state. We'll adopt the convenient convention that if S rules out A, then $S[A]$ is stipulated to be the **absurd state**, which we denote $\bot$. We also adopt the convention that $\bot[A] = \bot$ for any proposition A.

Similarly, A and B are **incompatible** (for a state S) iff $S[A][B] = \bot$. That is, either S rules out A, or else $S[A]$ is a reasonable state that rules out B. (This is analogous to the probabilistic definition of compatibility in Section 1.) Finally, A **implies** B for a state S iff A is incompatible with $\neg B$ for S. We should emphasize that even though these terms are obviously suggestive of similar notions involving sets of possible worlds, classical consequence, or probability, we don't mean to build in assumptions from any of those domains. We'll explicitly state the structural assumptions that our results rely on.

In this framework, assigning credence zero to a proposition no longer plays any important role. Indeed, assigning numerical credences to propositions no longer plays any role at all. So let's restate a pared down version of Stalnaker's Thesis that does without numerical credences as well.

> **New Stalnaker's Thesis.** Let S be any reasonable state that does not rule out A. Then
> $$S \text{ rules out } (\text{if } A, B) \quad \text{iff} \quad S[A] \text{ rules out } B$$

Notice the obvious way in which this is weaker than Stalnaker's Thesis: it only applies to an extreme attitude toward the conditional—ruling it out—and puts no direct constraints on intermediate attitudes.[12]

In Section 1 we showed that Stalnaker's Thesis leads to an unacceptable conclusion. We'll now extend that argument to New Stalnaker's Thesis, using the following two assumptions:

> **Negation.** For any reasonable state S, the propositions $\neg A$ and A are incompatible.

> **Commutativity.** For any reasonable state S and propositions A and B,
> $$S[A][B] = S[B][A]$$

Note that unlike the principle of Commutativity that we discussed in Section 2, this principle is not a theorem of the probability calculus, but rather a basic postulate. In due course we'll explore the possibility of dropping it. Of course, if you interpret states to simply be (C, V) pairs, and $S[A]$ to be (C_A, V_A), then the postulate follows from the corresponding theorem. But our use of the postulate does not assume or depend on these identifications.

It's instructive to compare this Commutativity principle with an alternative thesis we briefly discussed in Section 2: "if the relevant evidence and experience collected is the same, then the order of collection should not matter for the final degrees of belief" (Arntzenius 2003, 364). While it is related, this thesis differs from our postulate in two important ways. First, while Arntzenius's thesis concerns arbitrary

[12] There is also a subtle way in which New Stalnaker's Thesis might be slightly stronger than the original. Since ruling out a proposition is no longer identified with assigning numerical credence zero, one might have thought it was compatible with Stalnaker's Thesis to have both $C(\text{if } A, B) = 0$ and $C(B \mid A) = 0$, while *ruling out* just one of them. But while it might obey the letter, clearly this would be out of the spirit of the original proposal, which says that one takes the same attitude toward (if A, B) as one conditionally takes toward B, given A.

"evidence and experience", our Commutativity postulate specifically concerns updating on propositions. Of course, if updating on evidence and experience always amounts to updating on some proposition, then this difference of formulation is insignificant. But Richard Jeffrey notably rejects this connection: for example, if you are presented with visual experience of the color of a cloth by candlelight, he argues that "a change in the probability assignment is clearly called for, but … the change is not occasioned simply by learning of the truth of some proposition *E*" (1983 [1965], 165). Moreover, on a natural construal of Jeffrey's alternative account of updating, the order in which non-propositional evidence is received can make a difference to one's final probabilities (Field 1978; Domotor 1980; see also Lange 2000; Weisberg 2009). Thus Jeffrey's account is at odds with Arntzenius's thesis; but it is not directly at odds with Commutativity as we have construed it, since Jeffrey-updating is not a rule for updating on propositional evidence. (That said, Jeffrey's view of evidence does present a challenge to the generality of our framework, though we won't explore that challenge here.)

The second point of contrast is that Arntzenius's version of the thesis concerns the (presumably temporal) "order of collection" of evidence. But our Commutativity postulate does not explicitly involve an order in time, but only the "order of operations" in the mathematical sense. The state $S[A][B]$ is the result of updating with B the state that results from updating S with A. Still, there is a straightforward connection between "atemporal" Commutativity and temporally ordered updates, since the beings whose epistemic life we wish to describe using these abstract states typically realize them in a certain temporal order. You're in state S, and *then*, after supposing A, enter state $S[A]$, and then after supposing something further, end up in $S[A][B]$. Neither the states nor the update function themselves require this ordering. By analogy, the abstract definition of a transition function t for a Turing machine does not explicitly say anything temporal: but any natural physical realization of this abstract machine will first be in a state S, and then later in time be in the state $t(S)$.

Commutativity has two important corollaries we'll use. (In fact, the argument against New Stalnaker's Thesis only relies on these weaker corollaries, rather than full-fledged Commutativity.)

> **Symmetry.** For any reasonable state S, if A and B are incompatible, then B and A are incompatible.

(If A and B are incompatible for S, then by Commutativity $S[B][A] = S[A][B] = \bot$.) This in turn implies:

> **Monotonicity.** If S rules out A, then $S[B]$ rules out A.

(If $S[A] = \bot$ then $S[B][A] = S[A][B] = \bot[B] = \bot$.)

Now consider any reasonable state S such that A and $\neg B$ are compatible. Negation tells us that $S[\neg B]$ rules out B, and Monotonicity then implies that $S[\neg B][A]$ rules out B. So applying New Stalnaker's Thesis to $S[\neg B]$ tells us that $S[\neg B]$ rules out (if A, B). This means that $\neg B$ is incompatible with (if A, B); or by Symmetry, (if A, B) is incompatible with $\neg B$. In short, for any reasonable state for which A and $\neg B$ are compatible, (if A, B) implies B. This conclusion is exactly parallel to the conclusion that Lewis originally derived; as we discussed earlier, this conclusion is unacceptable.

Symmetry and Negation were essentially the only assumptions we needed to prove this result. So neither of the two maneuvers Lewis proposed are to the point: the proof never involves embedding a conditional (not even under negation), and it never appeals to additivity. If denying that conditionals have truth conditions is going to provide any way of retaining New Stalnaker's Thesis (a thesis which is at least as compelling as the old one), it will have to be elaborated some third way.

It's also worthwhile to consider how this result connects to results concerning a different "qualitative" version of Stalnaker's Thesis (see Gärdenfors 1986; Stalnaker 1976b; Harper 1976). Let's say S **accepts** a proposition A iff $S[A] = S$: adding A as a supposition makes no difference when A is already taken for granted.[13]

> **Harper's Thesis.** For any reasonable state S, if S does not rule out A, then
>
> $$S \text{ accepts (if } A, B) \text{ iff } S[A] \text{ accepts } B$$

This thesis corresponds to the special case of Stalnaker's Thesis for probability *one*, rather than zero (if we set aside parallel concerns about how probability one and acceptance might come apart in certain cases).

We can make similar trouble for Harper's Thesis if we add two further assumptions:

> **Idempotence.** For any reasonable state S and proposition A,
>
> $$S[A][A] = S[A]$$

[13] Introducing this notion of acceptance raises some questions about the interaction between it and the notion of implication, as we've defined it. In particular, this is a plausible connection:

> A implies B for S iff $S[A]$ accepts B.

This connection can be derived using a further reductio-like postulate:

> S rules out $\neg B$ iff S accepts B.

(Apply the postulate to the state $S[A]$ and use the definitions.) This postulate might be contested—for instance, by intuitionists—but it's worth noting that it is upheld by the standard dynamic model of negation we discuss in Section 7 (Veltman 1996, Def. 2.3).

Supposing something twice over is no different from just supposing it once. An alternative way of stating Idempotence is that $S[A]$ accepts A.

> **Conditional Non-Contradiction.** For any reasonable state that does not rule out A, (if A, $\neg B$) and (if A, B) are incompatible.

If the bottle might have fallen, and you suppose that if the bottle fell it didn't break, then you thereby rule out the proposition that if it fell it broke.

Given Commutativity, Idempotence, and Conditional Non-Contradiction, we can similarly conclude that Harper's Thesis implies that contingent conditionals imply their consequents. The proof is mostly the same as before. Suppose that S is any reasonable state for which A and $\neg B$ are compatible—so $S[\neg B]$ and $S[\neg B][A]$ are reasonable states. Then by Commutativity $S[\neg B][A] = S[A][\neg B]$, which accepts $\neg B$ by Idempotence. By Harper's Thesis, $S[\neg B]$ accepts (if A, $\neg B$). Then by Conditional Non-Contradiction, $S[\neg B]$ rules out (if A, B), and thus by Symmetry, (if A, B) implies B.[14]

(We should comment briefly here on how this relates to an earlier triviality result of Peter Gärdenfors (1986), in a similar update-centered framework. Gärdenfors claims that this result

> is an even more general version of Lewis's triviality result. It is much more general since it depends neither on the assumption that consistent revision should be made by conditionalization, nor on the assumption that states of belief should be modelled by probability functions. (1986, 89)

But closer inspection reveals that this isn't exactly right: in fact, Gärdenfors uses assumptions that conditionalization does not obey—and indeed, which are inconsistent with Commutativity. This is a key assumption of Gärdenfors's proof (proposition (2) on p. 84, translated into our terminology and notation):

> For any reasonable state S, if S rules out A, then A is logically inconsistent.

This might initially sound plausible: only the absurd leads to absurdity. But it implies in particular that S[snow is not white] does not rule out (snow is white): the

[14] Here's an alternative package of assumptions we could use to derive similar trouble for Harper's Thesis. Rather than defining acceptance in terms of update, we could take acceptance as primitive, and use these linking principles instead:

> If A and B are incompatible for S, and S accepts A, then S rules out B.
>
> If S rules out A, then S accepts $\neg A$.

(That is, A and $\neg A$ are *exhaustive* in a certain sense.) Using these assumptions together with Symmetry, Negation, and Conditional Non-Contradiction, we can again show that Harper's Thesis entails that contingent conditionals imply their consequents. The proof goes very similarly.

framework requires updating sequentially on $\neg A$ and A to have a well-defined non-absurd result, unless A is ruled out by any state whatsoever.[15] Gärdenfors also uses these (plausible) assumptions:

> $S[A]$ accepts A
>
> No reasonable state accepts both A and $\neg A$

Let A be neither a tautology nor a contradiction. Then according to these assumptions, $S[A][\neg A]$ accepts $\neg A$, $S[\neg A][A]$ accepts A, and both of these are reasonable states—and so they are distinct. So Gärdenfors's assumptions are inconsistent with Commutativity. But Lewis's result—and ours—concern *commutative* updates, like conditionalization. Thus, while there is a strong superficial similarity between Gärdenfors's result and ours, in fact they are doing importantly different things.)[16]

5 Streamlining the Other Arguments

Applying this framework, we can see that the arguments from Section 3 rely on similarly minimal assumptions. We'll begin with "might".

> **New Might.** For any reasonable state S that does not rule out $(\mathsf{might}\, A)$,
>
> $$S[\mathsf{might}\, A] \text{ does not rule out } A$$

(Under the simplifying assumption that ruling out A is the same as having zero credence in A, then Might and New Might say the same thing. But for the reasons we've discussed—recall the dartboard example—we don't assume that.)

As before, this principle has the bad consequence that $\mathsf{might}\, A$ implies A. Here's the argument. Suppose that S is a reasonable state for which $\neg A$ and $(\mathsf{might}\, A)$ are compatible. Then $S[\neg A]$ does not rule out $(\mathsf{might}\, A)$, and applying New Might to

[15] The thought here is that when you update on A, you then *give up* any previous supposition of $\neg A$. This is in the spirit of theories of "non-monotonic" updating—like AGM belief revision—where updates don't merely add information, but can also take it way. (Not coincidentally, Gärdenfors is the "G" in "AGM".) But standard conditionalization doesn't work like this.

[16] Stalnaker proves another related result (1976b, in correspondence with Harper; see Harper 1976). Its context is Popper functions, which build in substantial probabilistic assumptions. But Stalnaker's proof only makes trouble for the stronger thesis that $P(B \mid A) = 1$ iff $P(\text{if } A, B) = 1$ *even when* $P(A) = 0$. (Specifically, see step 5 on p. 114: note that the assumptions there guarantee that $P_{A \wedge X}(Y \vee Z \vee B) = 0$.) In contrast, our result applies even with no assumptions about what happens to a conditional when its antecedent is ruled out—as is fitting, since conditionals with epistemically impossible antecedents are a deviant case.

this state tells us that $S[\neg A][\text{might}\, A]$ does not rule out A. But by Negation $S[\neg A]$ rules out A, so by Monotonicity $S[\neg A][\text{might}\, A]$ does rule out A. This is a contradiction, so we can conclude that $\neg A$ and $(\text{might}\, A)$ are incompatible for any reasonable state S, which is to say (by Symmetry) that $(\text{might}\, A)$ implies A. Again, the only assumptions we needed for this argument were Negation and Symmetry.

We'll briefly point out how things go for the rest. For the theses about probability, we'll first need to say something about the connection between the abstract "states" of our framework and assignments of probability. Let's suppose that for a state S and a proposition A, $C_S(A)$ is the credence assigned to A in state S. To avoid excessive subscripts, we'll use $C_S(B \mid A)$ as an alternative way of writing $C_{S[A]}(B)$.

Here's a natural hypothesis:

> **Zero.** If S rules out A then $C_S(A) = 0$.

(Note that we won't really need credences to be on a numerical scale for this to make sense—we'll just need comparisons of lower or higher levels of confidence, and a *lowest* level of confidence which we can call "zero".)

Now we can state natural analogues to our original theses about probability.

> **New Probably.** For any reasonable state S that does not rule out $(\text{probably}\, A)$,
>
> $$C_S(A \mid \text{probably}\, A) > C_S(\neg A \mid \text{probably}\, A)$$
>
> **New Exact Probability.** For any reasonable state S that does not rule out (the probability of A is x),
>
> $$C_S(A \mid \text{the probability of } A \text{ is } x) = x$$
>
> **New Comparative Probability.** For any reasonable credal state S that does not rule out (A is more probable than B),
>
> $$C_S(A \mid A \text{ is more probable than } B) > C_S(B \mid A \text{ is more probable than } B)$$

The argument goes basically the same way in each case. For instance, suppose that there is a reasonable state S for which $\neg A$ is compatible with $(\text{probably}\, A)$. In that case, by New Probably, $C_S(A \mid \neg A, \text{probably}\, A)$ should be high, and thus non-zero, and thus $S[\neg A][\text{probably}\, A]$ should not rule out $\neg A$. But as before, by Negation and Monotonicity, $S[\neg A][\text{probably}\, A]$ does rule out A. So we conclude that $\neg A$ is incompatible with $(\text{probably}\, A)$, or (by Symmetry) $(\text{probably}\, A)$ implies A.

For the theses about value, we also need to say something about how the abstract "states" of our framework connect to value. Let's suppose that for a state S, there is a certain degree of desirability V_S, which is your subjective evaluation of the goodness of reality. Again, to avoid messy subscripts, we'll use $V_S(A)$ as an alternative way of writing $V_{S[A]}$, which is your evaluation for how things are conditional on A. (As before, note that $V_S(A)$ may not make sense when S rules out A.)

> **New Comparative Value.** Let S be any reasonable state which does not rule out (A is better than B). Let $S_+ = S[A \text{ is better than } B]$. Then if S_+ does not rule out either of A or B, $V_{S_+}(A) > V_{S_+}(B)$.
>
> **New Good News.** Let S be any reasonable state which does not rule out $\text{good}\,A$. Let $S_+ = S[\text{good}\,A]$. Then if S_+ does not rule out A, $V_{S_+}(A) > V_{S_+}$.
>
> **New Exact Value.** Let S be a reasonable state which does not rule out (the value of A is x). Let $S_+ = S[\text{the value of } A \text{ is } x]$. Then if S_+ does not rule out A, $V_{S_+}(A) = x$.

Given Commutativity and Idempotence, these theses lead to trouble. Take the case of New Comparative Value. Suppose for reductio that S is a state for which A, B, and (A is better than B) are compatible. In that case, we can apply New Comparative Value to $S[A][B]$, letting

$$S_+ = S[A][B][A \text{ is better than } B]$$

By Commutativity and Idempotence, $S_+[A] = S_+ = S_+[B]$. But this implies that $V_{S_+}(A) = V_{S_+}(B)$, which violates New Comparative Value. So New Comparative Value implies that A, B, and (A is better than B) are incompatible, for any reasonable state S; or by Symmetry, (A is better than B) is incompatible with A and B. As before, that conclusion is very weird.

The proof that New Good News implies Converse Pessimism ($\text{good}\,A$ implies $\neg A$) is similar.

In the case of New Exact Value, let x and y be distinct, let X be (the value of A is x), let Y be (the value of B is y), and suppose for reductio that A, B, X, and Y are jointly compatible. In that case we can apply New Exact Value to the state $S' = S[A][B][X][Y]$ twice:

$$V_{S'[X][A]} = V_{S'[X]}(A) = x$$
$$V_{S'[Y][B]} = V_{S'[Y]}(B) = y$$

But also, Commutativity and Idemopotence imply

$$\begin{aligned} S'[X][A] &= S[A][B][X][Y][X][A] \\ &= S[A][B][X][Y] \\ &= S[A][B][X][Y][Y][B] \\ &= S'[Y][B] \end{aligned}$$

Together these equations contradict the fact that $x \neq y$. So (the value of A is x), (the value of B is y), A, and B, are jointly incompatible. This conclusion is also unacceptable.[17]

The idea of all three of these proofs is the same: no news is not good news. If a state S accepts A, then $V_S(A)$, the value S assigns reality conditional on A, is just the value V_S that S assigns reality already. In particular, if S accepts both A and B, then S must assign each of them the same value.

6 Parameterizing the Principles

So far we've seen that a simplified version of Stalnaker's Thesis and a host of similarly plausible-sounding principles have unacceptable consequences given very lean assumptions about how to update with new information. Subsets of Commutativity, Idempotence, Negation, and Zero suffice for us to derive absurdities from the principles. (Specifically: Commutativity and Idempotence for the evaluative principles; Negation and Symmetry, which follows from Commutativity, for "might" and "if"; Negation, Symmetry, and Zero for the probability principles.) These derivations don't turn on any details of a theory of probability, or on embedding conditionals, modals, or evaluative claims in conjunctions or disjunctions. Even our use of Negation is innocent enough in this respect: we only ever needed to negate "vanilla" propositions involving no problematic conditionals, modals, or evaluations. So a restricted version of Negation, which says that $\neg A$ is incompatible with A *as long as* A is "vanilla", will serve our purposes just as well.

Given how natural these assumptions are, the pressure is strong to reject the original principles as based on some kind of mistake. The literature contains some standard suggestions about how this diagnosis might go. One standard thing to say is that conditionals, modals, and evaluatives involve an autobiographical element, so the

[17] As we noted, Lewis's original Desire as Belief thesis is a bit less natural than these. Deriving trouble for it in our framework also uses somewhat different assumptions. In particular, unlike the arguments against the other theses, the argument against Lewis's Desire as Belief does depend on embedding (good A) under disjunction (or, in alternative forms of the argument, conjunction or negation).

truth conditions for the sentence “it might be the case that A” vary along with the epistemic state of the speaker.[18] There is no single proposition perspicuously labeled (might A); rather, it would be more perspicuous to write (might$_S$ A), parameterized with a state S. In that case New Might should be rewritten as well.

> **Parameterized Might.** If S does not rule out might$_S$ A then S[might$_S$ A] does not rule out A.

This revised principle is not vulnerable to the triviality result. Recall that the result's proof involved applying New Might to the state $S[\neg A]$. Correctly parameterized, the conclusion of the argument is that might$_{S[\neg A]}$ A implies A. But it is plausible that, for any reasonable state, this “might” claim *should be ruled out*—in which case the implication would hold trivially. Suppose, as a simple toy model, that might$_S$ A is the proposition that that S does not rule out A. Then might$_{S[\neg A]}$ A would say that $S[\neg A]$ does not rule out A. Anyone who knew the fact Negation, that $S[\neg A]$ does rule out A, would then rule this out. So the properly parameterized conclusion of the argument is acceptable after all.

This diagnosis works the same way for each of the principles. For instance, suppose (A is better$_S$ than B) is the proposition that $V_S(A) > V_S(B)$. For any state that already includes the information that A, $V_S(A)$ is just the same as the value of a tautology. So for a state with both A and B added, $V_S(A) = V_S(B)$. Thus anyone who knows this much about how subjective value works should rule out (A is better$_{S[A][B]}$ than B). As in the case of “might”, then, the parameterized conclusion of the triviality argument turns out to be acceptable.

One thing worth noting is that if features of your own state are not always open to introspection, then even principles like Parameterized Might are vulnerable to a separate line of attack.[19] Consider again the toy model for (might$_S$ A), and suppose also that it's possible to be in a state S that rules out A, without ruling out that S does not rule out A.[20] (To see the structure here, note that if ruling out A were the

[18] Van Fraassen (1976) gives this reply to the original triviality results for conditionals. Hájek and Pettit (2004) defend this kind of reply to Lewis's “desire as belief” triviality result—they call it the “indexicality loophole”. Kratzer's (1977) semantics for epistemic (and other) modals explicitly involves a contextually variable “modal base”.

[19] In the special case of conditionals, there are further independent attacks on Parameterized Stalnaker's Thesis, from *non-dynamic* triviality results (see citations in footnote 1). These other results—which turn on either details of the logic of conditionals embedded in conditionals, or (in the case of Hájek 1989) on a semantics that assigns each sentence a finite set of worlds. For critical discussion see Bacon (2014, sec. 2.2). These other results don't generally have obvious parallels for the other theses we've discussed.

[20] Note that “rules out” is hyperintensional. If you're in the state $S[\neg A]$, which rules out A, then you *should* rule out that $S[\neg A]$ doesn't rule out A. But just because your state is $S_+ = S[\neg A]$, it doesn't

same thing as knowing $\neg A$, then this would amount to a failure of the KK principle, that if you know $\neg A$ then you know you know $\neg A$. For arguments for the existence of such cases see e.g. Williamson (2002 ch. 4). For dissent see e.g. Greco (2014).) In that case—since S does not rule out (S does not rule out A), and $\mathsf{might}_S A$ is the proposition that S does not rule out A—S doesn't rule out $\mathsf{might}_S A$. But also, since S rules out A—assuming Monotonicity—$S[\mathsf{might}_S A]$ still rules out A. This contradicts Parameterized Might.

Similar considerations apply even to the parameterized principles about probability and value. Take this one:

> **Parameterized Exact Value.** Let S be a reasonable state which does not rule out (the value$_S$ of A is x). Let $S_+ = S$[the value$_S$ of A is x]. Then if S_+ does not rule out A, $V_{S_+}(A) = x$.

Let's consider a toy model of how uncertainty about subjective value might work. Your state S consists of a set of worlds, as does each proposition. A state S rules out a proposition A iff their intersection is empty. You have a certain fixed assignment of a value $V(w)$ to each possible world w. (For simplicity, let $V(w)$ be a number.) The subjective value for S of the proposition A is the average value of the worlds that are in both S and A.[21] You can be in doubt about your own state, so for each world w in S, there is a certain state $S(w)$—your state according to that world—which may vary from world to world, and in particular need not be identical to S.

Here's a simple case (Figure 1). You're either looking at a real Rembrandt or an excellent forgery; if it's real, then you can rule out it being fake; but if it's an excellent forgery, you can't rule out it being real.[22] Keeping things simple, let A be the proposition that is true at just two possible worlds, Good—it's a Rembrandt—and Bad—it's a forgery. Your state in the Good world S(Good) rules out Bad but not Good, while S(Bad) doesn't rule out either world. Say $V(\text{Good}) = 15$ and $V(\text{Bad}) = 5$. Thus the subjective value for S(Good) is 15, and the subjective value for S(Bad) is 10, the average of V(Good) and V(Bad). Accordingly, the proposition (the value$_S$ of A is 15) is true just at Good, and the proposition (the value$_S$ of A is 10) is true just at Bad.

follow that you also have to rule out that S_+ rules out A. Even if your state includes $\neg A$, you might truly say "I don't know whether *this state* rules out A". By analogy: you might be the loser, and rule out the loser winning, without ruling out *you* winning.

[21] This is just the expected value, where we're assuming for simplicity that the prior world-probabilities are uniform.

[22] Cases like this are often discussed in connection to the asymmetry of epistemic accessibility. For example, it's often thought that being embodied is compatible with the evidence a brain in a vat has, but not vice versa (see e.g. Williamson 2002, ch. 7).

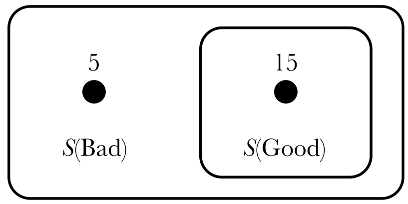

Figure 1: A case of uncertainty about one's own subjective values.

Suppose you are actually in the Bad world, so $S = S(\text{Bad}) = \{\text{Good}, \text{Bad}\}$ (even though you don't know this). In that case, you don't rule out the value$_S$ of A being 10. (Indeed, it *is* 10.) But consider

$$S_+ = S[\text{the value}_S \text{ of } A \text{ is } 10]$$

This adds to S the information about what your subjective value *in* S is. Note that when you learn this, you thereby learn what your state is. Learning that the value$_S$ of A is 10 eliminates the possibility that (the value$_S$ of A is 15) is true, and thus it eliminates the possibility that you are in the Good world. So S_+ is the singleton set containing just Bad. In that case, your new evaluation in S_+ is just the value of the Bad world, namely 5. So $V_{S_+}(A) = 5$, and not 10 after all. This contradicts Parameterized Exact Value.[23]

Aside from these technical objections, it's worth noting that the parameterized principles lose some of the initial appeal of the original formulations. After all (as two referees pressed on us) why should learning your *old* value constrain you to have that same value in your *new* state? Perhaps this constraint can be motivated, but it by no means has the same straightforward ring of plausibility as the simpler original version.

7 Truth Conditions and Commutativity

Earlier we considered and rejected two ways of elaborating the idea that conditionals (modals, evaluatives) lack truth conditions—either by saying that they don't embed under conjunction, or by giving up the additive law for probabilities. Saying either of those things can't help save the plausible-sounding principles, since our arguments raise trouble for those principles without relying on embedding or details about

[23]This case is analogous to Maria Lasonen-Aarnio's "Clock Belief" (2015, sec. IV; see also Christensen 2010). That example amounts to a counterexample to the parameterized version of the principle we've called Exact Probability.

probabilities. Now we're in a position to consider what a third elaboration of the "no truth value" position might look like. There's really only one assumption common to all of the arguments that is worth targeting: Commutativity.

Note that Commutativity naturally falls out of a standard truth-conditional picture, where adding information to an epistemic state works by shrinking a set of live possibilities, eliminating those possibilities at which the new information isn't true. If that's how things work, then Commutativity looks no stronger than a mundane fact about set intersection. If W is a set of live possibilities, and we update with A and then B, this leaves us with just the possibilities at which A is true and B is true—which are precisely the same as the possibilities at which B is true and also A is true. To give up Commutativity requires giving up this picture. So it wouldn't be surprising to see those who give up Commutativity also rejecting the "set of live possibilities" picture.

Indeed, this is what we find. Within formal semantics, the main rival to the standard truth-conditional picture is *dynamic semantics*, and one of its signal features that it permits violations of Commutativity.[24] The key thought is that, in general, the semantic value of a sentence is not a truth condition—something like a set of possibilities—but rather a "context change potential": a rule for updating the conversational state of play, or "context". While set intersection must commute, this is a special feature of that particular kind of update; there is no reason to think in advance that arbitrary rules for updating a context will commute. This just isn't how arbitrary sequences of changes work. Following the recipe, "Add two eggs, stir, and bake for 30 minutes" has a very different result from "Stir, bake for 30 minutes, and add two eggs."

Let's sketch a simple (and quite standard) model for how this goes for "might" (translating things along the way into the terminology we've used in this paper). A state S is a set of possible worlds. There is one "absurd" state, the empty set. A "proposition"— something that can be supposed—can in general be modeled by any function from states to states.[25] We get the result of updating S with A by function application: so $S[A]$ is just $A(S)$. A state S rules out A iff $S[A]$ is the absurd state.[26]

[24] For an overview see von Fintel and Gillies (2007), and works cited therein. Rothschild and Yalcin (2015) argue that violations of Commutativity and Idempotence are in fact the *defining* feature of dynamic semantics.

[25] The word "proposition" is often reserved for our "vanilla" propositions; but we're using the word in a very general sense throughout this paper.

[26] It's worth noting that while the formal model we're discussing comes from semantics, we're really deploying it for *epistemic* purposes, as a model of how epistemic states should be updated with certain kinds of propositions. As a referee pointed out to us, the dynamicists' distinctive account of the semantic values of English sentences using "if", "might", or "should" is, strictly speaking, a separate issue. The epistemology and the semantics fit together naturally, but one could in principle accept

For "vanilla" propositions, like "snow is white" or "Bea was born in 1985", the update function takes each state S to the intersection of S with some set of worlds—the set of worlds at which the proposition is true. For propositions like these, updates do commute. But for other cases, like epistemic modals, it doesn't work like this. The effect of (might A) on a state S is a *test*. The test is passed, and the result is S, just in case S does not rule out A. Otherwise the test is failed, and the result is the absurd state.

$$S[\text{might}\, A] = \begin{cases} \bot & \text{if } S \text{ rules out } A \\ S & \text{otherwise} \end{cases}$$

This is a model of New Might. By the definition, if S rules out A, then S rules out (might A). Contrapositively, if S does not rule out (might A), then S does not rule out A, and in that case by the definition $S[\text{might}\, A] = S$. So if S does not rule out (might A), then $S[\text{might}\, A]$ does not rule out A.

To show that this model violates Commutativity, consider a state S that doesn't rule out either of A or $\neg A$. Then $S[\text{might}\, A]$ passes the test, and thus is the same as S. So $S[\text{might}\, A][\neg A] = S[\neg A]$, which is not absurd. On the other hand, $S[\neg A]$ *does* rule out A. So $S[\neg A][\text{might}\, A]$ fails the test, and is absurd. In particular,

$$S[\neg A][\text{might}\, A] \neq S[\text{might}\, A][\neg A]$$

(cf. Veltman 1996, 223). This is not only a counterexample to Commutativity but also a counterexample to the weaker principle of Symmetry: $\neg A$ is incompatible with (might A) for the state S, but (might A) is compatible with $\neg A$. This also provides a counterexample to the even weaker principle of Monotonicity. The negation of (might A) is interpreted as the "inverse test": S passes the test $\neg$(might A) if and only if it fails the test (might A). Since S doesn't rule out A, S rules out $\neg$(might A). But $S[\neg A]$ is not absurd, and doesn't rule out $\neg$(might A) (cf. Veltman 1996, 230).

It's also worth noting that the dynamic account of (might A) also permits counterexamples to Idempotence (see Dorr and Hawthorne 2013). In the dynamic model, the conjunction (A and B) isn't understood on the model of set intersection, but rather as *function composition*. That is, updating on (A and B) has the same effect as first updating on A and then updating on B:

$$S[A \text{ and } B] = S[A][B]$$

(So for the reasons we just discussed, (A and B) need not be equivalent to (B and A).) Now consider the proposition

$$B = \text{might}\, A \text{ and } \neg A$$

this model for epistemic update, while denying that it is a good model of natural language meaning.

As before, let S be a state that doesn't rule out either A or $\neg A$. When we update S with B, we first apply the test (which it passes) and then update with $\neg A$. So

$$S[B] = S[\text{might}\, A \text{ and } \neg A] = S[\text{might}\, A][\neg A] = S[\neg A]$$

This is not the absurd state. Now update this with B a second time.

$$S[B][B] = S[\neg A][B] = S[\neg A][\text{might}\, A][\neg A]$$

This time, $S[\neg A]$ fails the "might" test. This means $S[B][B]$ is absurd, and so it is distinct from $S[B]$.

We can also provide bare-bones models for the other principles along the same lines as this model for New Might. While they are instructive, we don't claim that these models are realistic. Avoiding the triviality results is one thing; fleshing out a positive theory of conditionals, modals, and evaluatives that predicts the relevant principles is quite another.

In the case of conditionals, if instead of full-fledged Stalnaker's Thesis we are only concerned with the limiting cases corresponding to probability one and zero—what we called New Stalnaker's Thesis and Harper's Thesis—we can provide a simple dynamic model.[27] In the framework we have just described, where states are sets of worlds, we define the update on a conditional proposition as follows (where $A \supset B$ is the material conditional):

$$S[\text{if}\, A, B] = \begin{cases} \bot & \text{if } S[A] \text{ rules out } B \\ S & \text{if } S[A] \text{ accepts } B \\ S[A \supset B] & \text{otherwise} \end{cases}$$

That is, in the extreme cases where a state conditionally accepts or rejects B, we treat the conditional as a test; in the intermediate cases we treat it as a material

[27] Recall that for the fully general version of Stalnaker's Thesis for conditionals, there are other obstacles arising from "static" triviality results. The dynamicist has an easy reply to the finite-set-of-worlds argument of Hájek (1989), since this approach abandons the idea that a conditional has a set of worlds as its semantic value. As for the argument of Stalnaker (1976a): unlike either the Lewis-style or Hájek-style arguments, this one *does* turn on embedded conditionals. As we've discussed, giving up embedded conditionals won't help in the context that Lewis suggested it; but it may have an alternative motivation as a reply to Stalnaker-style "static" theorems. (But see Bacon 2014 for an alternative reply.)

Moreover, as a referee pointed out, there may be other distinctive motivations for rejecting embeddings which are internal to the dynamic approach. Standard denotational semantics doesn't give much guidance about how to interpret tests embedded within tests, or within more general context change potentials. Some progress on this project has been made, but one could also head off some of these challenges by rejecting the felicity of certain embeddings.

conditional. The key features of the material conditional we are using here are (i) if S rules out $A \supset B$, then $S[A]$ rules out B; and (ii) if S accepts $A \supset B$, then $S[A]$ accepts B.[28] This model violates Symmetry: $S[\neg B]$ will always rule out (if A, B), but if A and B are logically independent then $S[\text{if } A, B]$ need not rule out $\neg B$.

We can use a similar "test" trick to give models for the probabilistic and evaluative principles. An example should be enough to get the idea across. We can let the proposition (A is better than B) also be a test, defined as follows:[29]

$$S[A \text{ is better than } B] = \begin{cases} S & \text{if } V_S(A) > V_S(B) \\ \bot & \text{otherwise} \end{cases}$$

It's clear that this model satisfies New Comparative Value: if S does not rule out (A is better than B), then it must be that $V_S(A) > V_S(B)$, and $S_+ = S[A \text{ is better than } B] = S$. As with the model for "might", this model violates Idempotence and Commutativity, as well as Symmetry and Monotonicity (which, recall, are weaker than Commutativity). For instance, let S be a state with three equiprobable worlds with values 5, 10, and 15; let A be true at just the 10-world, and let B be true at just the 5 and 15-worlds. Then in S the expected value for A is no greater than the expected value for B (both are 10), and thus S rules out (A is better than B). But if C is only true at the 5-world and the 10-world, then $S[C]$ assigns value 10 to A and value 5 to B, and so $S[C]$ does not rule out (A is better than B). In these dynamic models, ruling out a proposition does not preclude ceasing to rule it out after an update. (Similar tests provide models for the other principles in Section 5.)

For anyone who wants to hang on to some or all of Stalnaker's Thesis and the other analogous theses we presented (in their original non-parametrized forms), the only real option is to go in for some account that violates Commutativity for conditionals, epistemic modals and evaluatives. (It's still open to maintain a restricted version of Commutativity that only applies when both propositions are "vanilla".) This will likely involve rejecting an orthodox truth-conditional approach. The most promising way forward for the stalwart defender of such theses isn't any of the usual expressivist maneuvers (rejecting embeddings or standard probability theory) but rather something like this dynamic approach.

Conversely, if you are convinced that your attitude conditional on supposing A and then supposing B is inevitably the same as your attitude conditional on supposing B

[28] There are other ways of achieving the same effect. For instance, the last clause of the definition could just as well use $A \wedge B$ in place of $A \supset B$.

[29] We are assuming in the background some function V that assigns a number V_S to each reasonable state S. This could be given in the standard way, as an expectation operator on the utilities of the worlds in S, but for our purposes it doesn't matter.

and then supposing A (even for conditionals, etc.), then you should despair of keeping the theses. In that case, you should seek to explain away their initial attraction. The points we noted about subtler parameterized versions of the principles, and failures of luminosity—which are easy to overlook—may help with this project.

Once we recognize the simplicity of the core arguments underlying Lewis's twin triviality theorems, and the ease with which they generalize, it becomes clear that there only a few defensible ways forward. This isn't the place to adjudicate between them.

References

Adams, E. W. 1975. *The Logic of Conditionals: An Application of Probability to Deductive Logic*. Springer Science & Business Media.

Arló Costa, Horacio Arló, John Collins, and Isaac Levi. 1995. "Desire-as-Belief Implies Opinionation or Indifference." *Analysis* 55 (1): 2–5. doi:10.2307/3328612.

Arntzenius, Frank. 2003. "Some Problems for Conditionalization and Reflection." *Journal of Philosophy* 100 (7): 356–70. http://www.jstor.org/stable/3655783.

Bacon, Andrew. 2014. "Stalnaker's Thesis in Context." *The Review of Symbolic Logic*, September, 1–33. doi:10.1017/S1755020314000318.

Bennett, Jonathan. 2003. *A Philosophical Guide to Conditionals*. Oxford University Press. http://dx.doi.org/10.1093/0199258872.001.0001.

Chalmers, David J., and Alan Hájek. 2007. "Ramsey + Moore = God." *Analysis* 67 (294): 170–72. http://www.jstor.org/stable/25597798.

Christensen, David. 2007. "Epistemic Self-Respect." *Proceedings of the Aristotelian Society*, New series, 107 (January): 319–37. http://www.jstor.org/stable/20619361.

———. 2010. "Rational Reflection." *Philosophical Perspectives* 24 (1): 121–40. doi:10.1111/j.1520-8583.2010.00187.x.

Domotor, Zoltan. 1980. "Probability Kinematics and Representation of Belief Change." *Philosophy of Science* 47 (3): 384–403. http://www.jstor.org/stable/186951.

Dorr, Cian, and John Hawthorne. 2013. "Embedding Epistemic Modals." *Mind* 122 (488): 867–913. doi:10.1093/mind/fzt091.

Edgington, Dorothy. 1986. "Do Conditionals Have Truth-Conditions." *Crítica* 18 (52): 3–30.

———. 2014. "Indicative Conditionals." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2014. http://plato.stanford.edu/archives/win2014/entries/conditionals/.

Field, Hartry. 1978. "A Note on Jeffrey Conditionalization." *Philosophy of Science* 45 (3): 361–67. http://www.jstor.org/stable/187023.

Fitch, Frederic B. 1963. "A Logical Analysis of Some Value Concepts." *Journal of Symbolic Logic* 28 (2): 135–42. http://www.jstor.org/stable/2271594.

Gärdenfors, Peter. 1986. "Belief Revisions and the Ramsey Test for Conditionals." *The Philosophical Review* 95 (1): 81. doi:10.2307/2185133.

Greco, Daniel. 2014. "Could KK Be OK?" *Journal of Philosophy* 111 (4): 169–97.

Harper, William L. 1976. "Ramsey Test Conditionals and Iterated Belief." In *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, edited by William L. Harper and C.A. Hooker, Vol. I: Foundations and Philosophy of Epistemic Applications of Probability Theory:117–35. The University of Western Ontario Series in Philosophy of Science. Boston: Dordrecht, Holland.

Hausman, Daniel M. 1995. "The Impossibility of Interpersonal Utility Comparisons." *Mind*, New series, 104 (415): 473–90. http://www.jstor.org/stable/2254638.

Hájek, Alan. 1989. "Probabilities of Conditionals: Revisited." *Journal of Philosophical Logic* 18 (4): 423–28. http://www.jstor.org/stable/30226421.

———. 2003. "What Conditional Probability Could Not Be." *Synthese* 137 (3): 273–323. http://www.jstor.org/stable/20118365.

Hájek, Alan, and N. Hall. 1994. "The Hypothesis of the Conditional Construal of Conditional Probability." In *Probability and Conditionals: Belief Revision and Rational Decision*, edited by Ellery Eells, Brian Skyrms, and Ernest W. Adams, 75. Cambridge University Press.

Hájek, Alan, and Philip Pettit. 2004. "Desire Beyond Belief." *Australasian Journal of Philosophy* 82 (1): 77–92. doi:10.1080/713659805.

Jeffrey, Richard. 1983 [1965]. *The Logic of Decision*. University of Chicago Press.

Kratzer, Angelika. 1977. "What 'Must' and 'Can' Must and Can Mean." *Linguistics and Philosophy* 1 (3): 337–55. doi:10.1007/BF00353453.

Lange, Marc. 2000. "Is Jeffrey Conditionalization Defective by Virtue of Being Non-Commutative? Remarks on the Sameness of Sensory Experiences." *Synthese* 123 (3): 393–403. http://www.jstor.org/stable/20118290.

Lasonen-Aarnio, Maria. 2015. "'I'm onto Something!' Learning About the World by Learning What I Think About It." *Analytic Philosophy* 56 (4): 267–97. doi:10.1111/phib.12069.

Lewis, David. 1976. "Probabilities of Conditionals and Conditional Probabilities." *The Philosophical Review* 85 (3): 297–315. doi:10.2307/2184045.

———. 1986. "Probabilities of Conditionals and Conditional Probabilities II." *Philosophical Review* 95: 581–89. http://www.jstor.org/stable/2184045.

———. 1988. "Desire as Belief." *Mind* 97 (387): 323–32. doi:10.2307/2255078.

———. 1996. "Desire as Belief II." *Mind* 105 (418): 303–13. doi:10.2307/2254563.

Ramsey, Frank Plumpton. 1990 [1929]. "General Propositions and Causality." In *Philosophical Papers*, edited by D. H. Mellor, 145–63. Cambridge: Cambridge University Press.

Rothschild, Daniel, and Seth Yalcin. 2015. "On the Dynamics of Conversation." *Noûs*, September, n/a–/a. doi:10.1111/nous.12121.

Stalnaker, Robert. 1976a. "Letter to Harper." In *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, edited by William L. Harper and C.A. Hooker, Vol. I: Foundations and Philosophy of Epistemic Applications of Probability Theory:113–15. The University of Western Ontario Series in Philosophy of Science. Boston: Dordrecht, Holland.

———. 1976b. "Letter to van Fraassen." In *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, edited by William L. Harper and C.A. Hooker, Vol. I: Foundations and Philosophy of Epistemic Applications of Probability Theory:302–6. The University of Western Ontario Series in Philosophy of Science. Boston: Dordrecht, Holland.

Van Fraassen, Bas C. 1976. "Probabilities of Conditionals." In *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, edited by William L. Harper and C.A. Hooker. Vol. Vol. I: Foundations and Philosophy of Epistemic Applications of Probability Theory. The University of Western Ontario Series in Philosophy of Science. Boston: Dordrecht, Holland.

van Roojen, Mark. 2014. "Moral Cognitivism Vs. Non-Cognitivism." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Fall

2014. http://plato.stanford.edu/archives/fall2014/entries/moral-cognitivism/.

Veltman, Frank. 1996. "Defaults in Update Semantics." *Journal of Philosophical Logic* 25 (3): 221–61. http://www.jstor.org/stable/30226572.

von Fintel, Kai, and Anthony S. Gillies. 2007. "An Opinionated Guide to Epistemic Modality." In *Oxford Studies in Epistemology 2*, edited by Tamar Gendler and John Hawthorne, 2:32–62. Oxford.

Weisberg, Jonathan. 2009. "Commutativity or Holism? A Dilemma for Conditionalizers." *The British Journal for the Philosophy of Science* 60 (4): 793–812. doi:10.1093/bjps/axp007.

Williamson, Timothy. 2002. *Knowledge and Its Limits*. Oxford University Press. http://dx.doi.org/10.1093/019925656X.001.0001.

Yalcin, Seth. 2007. "Epistemic Modals." *Mind* 116 (464): 983–1026. doi:10.1093/mind/fzm983.

———. 2012. "Bayesian Expressivism." *Proceedings of the Aristotelian Society* 112 (2): 123–60. doi:10.1111/j.1467-9264.2012.00329.x.