Lecture Notes:
Probability of Conditionals in Modal Semantics

Justin Khoo and Paolo Santorio

November 4, 2018

Contents

Overview

These notes explore the relationship between natural language conditionals and probability, with particular emphasis on theories of conditionals in contemporary modal semantics. One central claim we investigate is the conjecture (sometimes called 'The Thesis') that the probabilities of conditionals are the conditional probabilities of the consequent given the antecedent. But the overarching theme that we are interested in is the relationship between modal notions in natural language and probabilistic attitudes towards them.

Disclaimer. (November 4, 2018) These lecture notes are a work in progress and are being freshly written as the course proceeds. Typos, incongruencies, and mistakes of various sorts are to be expected. Please let us know if you spot any!

Chapter 1

Foundations of modal and conditional semantics and Stalnaker's Thesis

Formal Preliminaries

- Let $\mathcal{W}$ be a set of worlds, and $\mathcal{F}$ be a set of subsets of $\mathcal{W}$ that is closed under Boolean operations (including countable union).
- Let $[\![\cdot]\!]$ be an interpretation function, which maps expressions to extensions relative to a context c and index i.
  - A context is a centered world, a triple of a world w, time t, and location l.
  - An index is a tuple of various parameters, including a world (we will add more parameters as we go along).
  - Let $i[w]$ be the index just like i except possibly that its world parameter is w. We can then say that the proposition expressed by a sentence A at c is:

    $$[\![A]\!]^c = \{w\colon [\![A]\!]^{c,i_c[w]} = 1\ \}$$

- Thus, $[\![\cdot]\!]$ maps sentences to propositions in $\mathcal{F}$ (perhaps relative to a context).
- Let P be a probability function over $\mathcal{F}$, and Pr a function on sentences in our target language. For now, we define Pr in terms of P as follows:

  $$Pr(A) = P(\mathbf{A});\ \text{where } \mathbf{A} = \{w\colon [\![A]\!]^{c,i_c[w]} = 1\}$$

  We suppress relativization to context to ease readability

We adopt the following typographical conventions:

- We use uppercase italic letters for sentences: $A, B, C, \ldots$; shorthand:
  - $AB = A \wedge B$
  - $\bar{A} = \neg A$

- We use uppercase bold letters for propositions: **A**, **B**, **C**, ...
- For any sentences A, B: $A \to B$ is the indicative conditional whose antecedent is A and consequent is B.

1.1 Beginnings: the material conditional

The material conditional $\supset$ is a two-place truth functional operator, defined by the following truth table:

A	B	$A \supset B$
T	T	T
T	F	F
F	T	T
F	F	T

Here are a few important properties of the material conditional, which distinguish it from natural language indicative conditionals:

- $A \supset B$ is true if A is false.
  - Intuitively, the falsity of A is not sufficient for the truth of $A > B$:

  (1) If Paolo is not one of the instructors of this course, then Paolo is American.

  - (1) is false, since there is no obvious relation between being American and not being an instructor of this course. Yet, its material cousin is true simply because it is false that Paolo is not an instructor of this course.
- $A \supset B$ is true if B is true.
  - Intuitively, the truth of B is not sufficient for the truth of $A > B$:

  (2) If Justin was born in the UK, then Justin is an American citizen.

  - (2) is false, since again there is no obvious relation between being born in the UK and being an American citizen. Yet its material cousin is true simply because it is true that Justin is an American citizen.
- $A \supset B$ is false only if A is true and B is false.
  - Intuitively, the falsity of $A \supset B$ does not require the truth of A:

  (3) If Sue was born in San Francisco, then Sue was born in Connecticut.

- (3) is false, regardless of where Sue was born, since SF is not in CT. So, its falsity does not entail that Sue was born in SF.

For our purposes, perhaps the most important difference between the material conditional and indicative conditionals is their probabilities. Suppose a fair six-sided die was rolled and the result kept hidden. Consider:

(4) If the die landed on a prime, it landed on an odd.

- What is the probability of (4)? If it were a material conditional, its probability would be 5/6—this is the probability that either it landed on a non-prime (1, 4, 6) or it landed on an odd (1, 3, 5). But intuitively its probability is lower, only 2/3.
- Why 2/3? The natural answer is because there are three prime outcomes (2, 3, 5) and two of them are odd (3, 5). We'll come back to this answer shortly.

1.2 Strict and variably strict conditionals

Basis of modal semantics. All modal semantics for conditionals descend from the semantics of modal logics. The basic elements of the model theory of a modal semantics are two:

- A domain $\mathcal{W}$ of possible worlds;
- an accessibility relation $\mathcal{R}$, that relates two worlds w and w' just in case w can 'see' w' for the purposes of evaluating modal claims.

Here are some basic semantic clauses for modal operators:

(5) a. $[\![\Box A]\!]^w = 1$ iff $\forall w' \in \mathcal{W} : w'\mathcal{R}w$, $[\![A]\!]^{w'} = 1$
 b. $[\![\Diamond A]\!]^w = 1$ iff $\exists w' \in \mathcal{W} : w'\mathcal{R}w$, $[\![A]\!]^{w'} = 1$

Rather than using talk of an accessibility relations, we will follow custom in the semantics literature and relativize interpretation to a **sphere of accessibility** (Lewis) or **modal base** (Kratzer). We take this to be just a function from a world to a *set* of accessible worlds, and (for now) represent it as a subscript 'f' on the modal.

(6) a. $[\![\Box_f A]\!]^w = 1$ iff $\forall w' \in f(w)$: $[\![A]\!]^{w'} = 1$
 b. $[\![\Diamond_f A]\!]^w = 1$ iff $\exists w' \in f(w)$: $[\![A]\!]^{w'} = 1$

The semantic clauses in (6) can provide the basic scaffolding for the semantics of modals. In particular, they can be used to give a semantics for modals irrespective of modal flavor. For example, consider epistemic and the deontic uses of *must* (exemplified, respectively, in (7a) and (7b)).

(7) a. Frida must be in Chicago.

b. Frida must go to Chicago.

Both the occurrences of *must* in (7) can be given a unitary analysis, which closely follows (6). The difference in modal flavor is taken to be due to a difference in contextual input: context selects a set of epistemically accessible worlds in one case, and deontically accessible worlds in another.[1] Since the value of f is going to be provided by context and can change from utterance to utterance, we are going to include f in the index of evaluation (rather than just having it be part of the background model). Here is an entry for *must*:

(8) $[\![\text{must}_f A]\!]^{w,f} = 1$ iff $\forall w' \in f(w)$: $[\![A]\!]^{w',f} = 1$

To avoid clutter, from now on we will omit the modal base subscript from the modal.

Strict conditional semantics. A **strict conditional** is a conditional that is equivalent to the result of prefacing a necessity modal in front of a material conditional:

$$\Box_f(A \supset C)$$

A first, natural analysis treats conditionals in natural language as strict conditionals. On this analysis, a conditional $A \rightarrow C$ has the schematic truth conditions:

(9) $[\![A \rightarrow C]\!]^{w,f} = 1$ iff $\forall w' \in f(w)$ s.t. $[\![A]\!]^{w',f} = 1$, $[\![C]\!]^{w',f} = 1$

We can assume that bare indicative conditionals, i.e. conditionals with no overt modal, default to having epistemic flavor. For the time being, let's assume that epistemic modal bases map each world to the set of worlds that are compatible with what is known in that world (leaving it open what the subject of the relevant body of knowledge should be). Then the truth conditions of a conditional in (10a) are in (10b):

(10) a. If Frida danced, Maria danced.
b. $[\![(10a)]\!]^{w',f} = 1$ iff all worlds w' compatible with what is known in w such that Frida danced in w' are such that Maria danced in w'

Variably strict conditional semantics. One prediction of a simple strict conditional analysis is that conditionals are monotonic in the restrictor position: a conditional entails all conditionals with stronger antecedents and identical consequents.

Antecedent strengthening $\quad A \rightarrow C \vDash A^+ \rightarrow C \quad$ (where $A^+ \vDash A$)

This prediction is clearly problematic for some kinds of conditionals. For example, it is problematic for counterfactuals. Consider the following discourse:

[1] We should emphasize that this is not the final form of the analysis. Especially for the case of deontic modality, we will need to make use of an ordering; for relevant arguments, see Kratzer 2012.

(11) If Otto had come, it would have been a lovely party.
If both Otto and Anna had come, it would have been a dreary party.

(11) is a **Sobel sequence**, i.e. a discourse of the apparent form

$$A \to C, A^{+} \to \neg C$$

Sobel sequences are predicted to be straightforwardly inconsistent in a strict semantics. Yet discourses like (11) sound consistent. This style of argument (among others) has pushed theorists of counterfactuals towards a nonmonotonic analysis.

The empirical situation is less clear for the case of indicative conditionals. Consider a Sobel sequence involving indicative conditionals, like (12):

(12) If Otto was at the party, the party was great.
If Otto and Anna were at the party, the party was not great.

It is unclear whether (12) is felicitous, at least if we understand it as a unified discourse involving no changes of mind or corrections on the part of the speaker. Some (in particular, Williams 2008) take it to be a genuine Sobel sequence, on a par with (11). More recently, Peter Klecha (2015) has argued that (12) is not a genuine Sobel sequence, since any coherent reading of it must involve a context shift between the two conditionals.

But there is a different, and much more direct route to argue for a nonmonotonic semantics if we're interested in the connection between probabilities of conditionals and conditional probabilities. Conditional probabilities are nonmonotonic, i.e. we can have both $Pr(C \mid A) < Pr(C \mid (A \wedge B))$ and $Pr(C \mid A) > Pr(C \mid (A \wedge B))$, depending on our choice of A, B, and C. For a simple illustration, suppose that we are throwing a fair die. Using numbers to represent propositions about the outcome in the obvious way, we have both:

$$Pr(5 \vee 6 \mid 3 \vee 4 \vee 5 \vee 6) < Pr(5 \vee 6 \mid 4 \vee 5 \vee 6)$$
$$Pr(5 \vee 6 \mid 3 \vee 4 \vee 5 \vee 6) > Pr(5 \vee 6 \mid 3 \vee 4 \vee 5)$$

I.e., the proposition '$3 \vee 4 \vee 5 \vee 6$' can be strengthened in two different ways, which pushes the relevant conditional probabilities in different directions.

On a monotonic semantics $A \to C$ entails $A^{+} \to C$, hence we invariably $Pr(A \to C) \leq Pr(A^{+} \to C)$. So, if we're interested in vindicating a link between conditional probabilities and probabilities of conditionals, a nonmonotonic semantics seems a better bet.

One standard way to move to a variably strict semantics is to introduce a notion of comparative closeness between worlds. Comparative closeness is a three-argument relation: we compare two worlds w' and w'' with regard to their closeness to a 'base' world w. Formally, this can be captured by introducing a (partial or total) preorder $\preceq_w$ on the worlds in the domain, and relativizing interpretation to a choice of preorder[2] (in addition to other parameters). We then

[2]Reminder: a preorder is a binary relation that is reflexive and transitive.

use $\preceq_w$ to single out a set of worlds that are closest to w, where a world counts as closest to w just in case there are no worlds that are closer to it.[3]

On the revised semantics, conditionals quantify over the closest worlds to w where the antecedent is true (which we represent as $\mathsf{max}_{\preceq_w}$).

(13) $[\![A \to C]\!]^{w,f,\preceq} = 1$ iff:
$\forall w' \in \mathsf{max}_{\preceq_w}(f(w))$ s.t. $[\![A]\!]^{w',f,\preceq} = 1$, $[\![C]\!]^{w',f,\preceq} = 1$

For our purposes, it's useful to focus on one particular variant of the semantics in (13), namely the one that uses selection functions. Generally, a *selection function* is a function $s : W \times \mathcal{P}(W) \mapsto \mathcal{P}(W)$ that maps a pair of a world and a set of worlds to a set of worlds. One case that is of particular interest to us is that where selection functions select a singleton (or, equivalently, just a world). Call a function of this kind a **Stalnaker selection function**.

Stalnaker selection functions can be used in a semantics for conditionals that has important properties when it comes to vindicating the relationship with probability. The basic idea behind this semantics is the idea that the closeness relation induces a linear order on $\mathcal{W}$. I.e., for every pair of worlds w, w' in $\mathcal{W}$, exactly one between $w \prec w'$ and $w' \prec w$ is true. As a result, for every conditional antecedent, there is exactly one closest world that makes that antecedent true.

Here is a formal characterization of a semantics based on Stalnaker selection functions. A **Stalnaker model** includes a set of possible worlds $\mathcal{W}^-$ as well as an *absurd world* λ, which makes true every sentence. Call the union of $\mathcal{W}^-$ and $\{\lambda\}$ simply $\mathcal{W}$. We define a Stalnaker selection function as follows.

A function $s : \mathcal{W} \times \mathcal{P}(\mathcal{W}) \mapsto \mathcal{W}$ is a **Stalnaker selection function** iff

i. if $[\![A]\!]$ is non-empty, $s(w, A) \in [\![A]\!]$
(**Inclusion:** the selected world must make true the antecedent, if at all possible.)

ii. if $s(w, A) = \lambda$, then $[\![A]\!] = \varnothing$ (where λ is the absurd world, i.e. a world where every sentence is true)
(**Absurdity-as-last-resort**: the impossible world is selected only if no possible world can be selected.)

iii. if $w \in [\![A]\!]$, then $s(w, A) = w$
(**Centering**: if the world of evaluation makes the antecedent true, it is identical to the selected world.)

iv. for all A, A': if $s(w, A) \in [\![A']\!]$ and $s(w, A') \in [\![A]\!]$, then:
$s(w, A) \in [\![A']\!] = s(w, A') \in [\![A]\!]$
(**Consistency of selection**: the selection must be consistent for all choice of antecedents.)

Below is a semantics of conditional based on Stalnaker selection functions. We relativize interpretation a Stalnaker selection function parameter s, in addition to the usual ones.

[3]Formally: $w \in \mathsf{max}_{\preceq_w})$ iff $\neg\exists w' : w' \preceq w$. Here and in our statement of truth conditions we are assuming the so-called limit assumption, i.e. the assumption that, for every antecedent A, and every world w, there is a set S of closest A-verifying worlds to w.

(14) $[\![A \to C]\!]^{w,f,s} = 1$ iff $[\![C]\!]^{s(w,f(w)\cap\mathbf{A}),f,s} = 1$

In less compressed form: $A \to C$ is true relative to w, f, and s iff the closest A-world, i.e. the world selected by s taking as input w and the proposition resulting from intersecting the modal base and **A**, is a C-world.

1.3 Restrictor Theory

So far, we have focused on simple conditionals (those with non-conditional and non-modal antecedents and consequents), and treated *if* as a two-place modal operator: $\to$. However, thinking about a wider range of conditionals may force us to consider alternative compositional theories. In particular, it has been argued that the interaction between *if*s and adverbial quantifiers, modals, and other *if*s motivates thinking about *if* not as a modal operator, but rather as a device for restricting the domains of other operators (cf. Lewis (1975); Kratzer (1981, 1986)). Call such a theory a **restrictor theory** of conditionals.

To see how a restrictor theory differs from an operator theory, recall the strict conditional theory from above:

Strict Conditional Theory
$[\![A \to C]\!]^{w,f} = 1$ iff $\forall w' \in f(w)$ s.t. $[\![A]\!]^{w',f} = 1$, $[\![C]\!]^{w',f} = 1$

Note that, here, $\to$ is shifting the world of evaluation of the consequent B to the worlds in the domain $f(w)$ at which A is true. However, notice that the consequent B is still evaluated relative to the same domain function f, just at a different world w'. Because of this, f may "forget" that we are only looking at A-worlds when we check for B's truth at w', f. As a result, the **Strict Conditional Theory** predicts that conditionals like (15) may fail to be true:

(15) If it rained, then if John brought his umbrella, it rained.

This seems wrong, though: (15) seems like it is equivalent to:

(16) If it rained and John brought his umbrella, it rained.

and thus necessarily true. Consider another example, this one involving a modal in the consequent of a conditional:

(17) If Jane drew a diamond, she must have drawn a red card.

Suppose we do not know what card Jane drew (and also that we know that we don't know this). Still, it seems that 17 is something we know to be true, simply because we know all the diamonds are red cards in a normal deck of cards. However, notice that the **Strict Conditional Theory** predicts that 17 is in fact false in this scenario. To see why, consider the truth conditions it assigns to 17:

(18) $[\![D \to \Box R]\!]^{w,f} = 1$ iff $\forall w' \in f(w)$ s.t. $[\![D]\!]^{w',f} = 1$, $[\![\Box R]\!]^{w',f} = 1$

Given that we know that we don't know what card drew, there are R and $\neg R$-worlds in the domain $f(w)$, and also for any $w' \in f(w)$: there are R and $\neg R$-worlds in $f(w')$ as well. Thus, it follows that $[\![\Box R]\!]^{w',f} = 0$, for any $w' \in f(w)$, and hence that $[\![D \to \Box R]\!]^{w,f} = 0$.

These problems arise for our operator implementation of the strict conditional theory because the operator $\to$ shifts only the world of evaluation, and does not shift the modal base f as well. If instead of evaluating $\Box R$ at f, w', we evaluated it at $f^{\mathbf{D}}, w'$, where $f^{\mathbf{D}}$ contains the information that Jane drew a diamond, we would correctly predict that (17) is true. Let's see why by first defining $f^{\mathbf{X}}$:

$$\text{For any } w : f^{\mathbf{X}}(w) =_{\mathbf{def}} f(w) \cap \mathbf{X}$$

Now, we can see that $[\![\Box R]\!]^{w',f^{\mathbf{D}}} = 1$. This is because we know that all of the diamonds are red cards (and we know that we know this). So, every $w'' \in f^{\mathbf{D}}(w')$ is such that $[\![R]\!]^{w'',f^{\mathbf{D}}} = 1$, which is sufficient for $[\![\Box R]\!]^{w',f^{\mathbf{D}}} = 1$.

We can now define a shifty strict conditional theory as follows:

Shifty Strict Conditional Theory
$[\![A \to B]\!]^{w,f} = 1$ iff $\forall w' \in f^{\mathbf{A}}(w)$: $[\![B]\!]^{w',f^{\mathbf{A}}} = 1$

This semantic theory just is the one defended by Gillies (2009, 2010). Gillies calls it a "doubly shifty" theory since the conditional operator shifts both the world of evaluation and the modal base. Given the assumption that epistemic modal bases are **Closed**, this theory makes several important predictions:

Closed
For any w, w': if $w' \in f(w)$ then $f(w) = f(w')$.

Predictions:

1. $(A \to \Diamond B) \equiv \Diamond(A \wedge B)$
2. $(A \to (B \to C)) \equiv ((A \wedge B) \to C)$

We are not quite yet to a restrictor theory, as originally defined by Lewis and Kratzer. The key move now is to drop the world shifting component of $\to$, leaving just the modal base shifting component. With that, we can now define a restrictor theory as follows (cf. Kratzer (1981, 1991)):

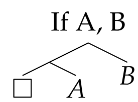
Restrictor Theory
$[\![A \to B]\!]^{w,f} = 1$ iff $[\![B]\!]^{w,f^{\mathbf{A}}} = 1$

Two observations about this theory:

1. If there is no modal in the consequent, then $A \to B$ ends up equivalent to B! To handle this issue, Kratzer proposes that bare conditionals (those without an overt modal) contain a covert, unpronounced, modal operator in their consequents. (This is not the only solution: one could instead opt for ambiguity, as Lewis apparently did. See Lewis (1975, 1976).)

(19) Notice that, if you add a covert $\Box$ in the consequent, then the restrictor theory predicts strict conditional truth conditions for *if A, B*.

2. It is sometimes said that the restrictor theory is a **syntactic** theory about *if*: that there is "no 'if ... then' connective in the logical forms for natural languages." (Kratzer (1986): 11). On that approach, conditionals in natural language have a special logical form, as follows:



However, we can now see that this syntactic formulation of the restrictor theory is inessential: **Restrictor Theory** as defined above makes all the predictions of the syntactic implementation discussed in Kratzer (1986). Furthermore, there are reasons to prefer the connective implementation to the syntactic implementation, for how else to handle conditionals like the following (cf. Gillies (2010))?

(20) If Jane drew a diamond, then she must have drawn a red card and she might have drawn an ace.

The reason we discuss the restrictor theory in a course on probabilities and conditionals is that the theory offers some strategies for resisting the various triviality proofs we will encounter. More on this in Chapter 3!

1.4 Stalnaker's Thesis

The Thesis, in full generality. Stalnaker's Thesis is a claim about the link between probabilities of conditionals and conditional probabilities.

Stalnaker's Thesis
For all Pr that model rational credence, and for all A, B:

$$Pr(A \to B) = Pr(B|A) \quad (\text{if } Pr(A) > 0)$$

For brevity (and to echo another label that is frequently used for it) we will frequently call it "the Thesis".

Notice two points. First, the Thesis is a *normative* rather than a descriptive claim about the probabilities of conditionals. It is consistent with Stalnaker's Thesis that there are subjects whose credences in conditionals fail to align with the relevant conditional probabilities. These subjects are ruled to be irrational, not impossible. Second, endorsing the Thesis entails assuming that there are rational constraints on credence that go beyond mere coherence constraints. Hence endorsing the Thesis means endorsing a kind of objective Bayesianism (though perhaps not the most common or frequently endorsed kind).

There are a number of strong considerations that favor the Thesis.

First of all, ordinary judgments about probabilities of conditionals very often align with the Thesis. For a simple example, suppose that Maria might have tossed a fair die, and assess the probability of (21):

(21) If Maria tossed the Die, it landed on 1 or 2.

The natural answer is '1/3', which of course is also the value of the corresponding conditional probability.

Second, the Thesis can be derived from two plausible principles about probabilities of conditionals. These principles are:

i. **Probabilistic Centering.** For all Pr that model rational credence, and for all A, B:

$$Pr(A \to B \wedge A) = Pr(A \wedge B)$$

Probabilistic Centering immediately follows from two widely accepted principles of conditional logic:

Strong Centering. $A \wedge B \vDash A \to C$
Weak Centering. $A \to C \vDash A \supset B$

Together, Strong and Weak Centering entail the equivalence of $A \to B \wedge A$ and $A \wedge B$ (and hence, *a fortiori*, their probabilistic equivalence). Notice that Weak Centering is the principle underlying the validity of Modus Ponens for natural language conditionals.

ii. **Independence.** The probability of a conditional is independent of the probability of its antecedent.

Independence
For all Pr that model rational credence, and for all A, B:

$$Pr(A \to B) = Pr(A \to B \mid A) \quad \text{(if } Pr(A) > 0\text{)}$$

The argument for Independence is, again,an argument from the plausibility of cases. (For a potential counterexample, see Rothschild 2013.)

Once we have Centering and Independence, we can give a quick proof of Stalnaker's Thesis:

i. $Pr(A \to B) =$

ii. $Pr(A \to B \mid A) =$ (via Independence)

iii. $\dfrac{Pr(A \to B \wedge A)}{Pr(A)} =$ (def of conditional probability)

iv. $\dfrac{Pr(A \wedge B)}{Pr(A)} =$ (Probabilistic Centering)

v. $Pr(B \mid A)$ (def of conditional probability)

Let us also notice that there are also entailments going in the other direction. Any two of Stalnaker's Thesis, Independence, and Probabilistic Centering allow us to derive the third. (Proofs left to the reader.)

Before Triviality: trouble from Modus Ponens? In the next chapter, we will explore in detail a large number of triviality results that put pressure on the tenability of the Thesis. Before that, let us point out that the interderivability of the Thesis, Probabilistic Centering, and Independence can potentially cause trouble. While Modus Ponens is unrestrictedly valid on traditional semantics for conditionals, this is not the case on other accounts. In particular, one of the hallmarks of restrictor semantics is that it invalidates Modus Ponens for the case of right nested conditionals, i.e. conditionals of the form ⌜If A, then, if B, then C⌝.[4] (Whether this happens will also depend on exactly what we take Modus Ponens to be; see below.) Hence we might worry that a failure of Centering might generate a failure of the Thesis.

To see this, let us use an example that is structurally analogous to McGee's famous counterexample to Modus Ponens (1985). Suppose that Maria tossed a fair, six-sided Die. Consider the following conditional.

(22) If the die landed even, then, if it didn't land on 2 or 4, it landed on 6.

(22) seems certain in any scenario where a fair, six-sided die is tossed. So we assume that one should assign to it credence 1, or near-1. But now, consider what credence one should assign to the conditional nested within (22) in isolation:

(23) If the die didn't land on 2 or 4, it landed on 6.

Intuitively, (23) should get credence 1/4. (Even if you want to resist assigning (23) a precise value, it seems clear that one should assign to it less than 1/2 credence, which is sufficient for the point below.)

Assuming that the claim that the die landed even gets probability 1/2, this entails that the conditional probability of the consequent of (22), given the antecedent, is different from 1. To see this, notice that (23) entails, via Weak Centering, that the die landed even. Hence we have:

$$\text{Pr(If not 2 or 4, 6} \wedge \text{even) = Pr(If not 2 or 4, 6)}$$

Hence the probability of (23), conditional on the die landing even, is 1/2.

(24) $$Pr(\text{If not 2 or 4, 6} \mid \text{even}) = \frac{Pr(\text{If not 2 or 4, 6} \wedge \text{even})}{Pr(\text{even})} = \frac{1/4}{1/2} = 1/2$$

Of course, one might respond in a variety of ways. (In particular, if one adopts the restrictor theory we presented in §1.3, it is not immediately clear what forms of argument we should take to be instances of Modus Ponens when nested conditionals are involved.) But the following points should be uncontroversial:

[4] See Khoo 2013 for discussion.

i. On any theory that treats conditionals as binary connectives, the example above amounts to a probabilistic failure of Modus Ponens.[5]

ii. Hence, again if we treat conditionals as binary connectives, we can observe a failure of Stalnaker's Thesis.

For this reason, alongside the unrestricted version of the Thesis, it's worth considering a restricted version of it, which only applies to simple conditionals (i.e., as we said in §1.3, conditionals with nonmodal and nonconditional antecedents and consequents.)

Restricted Stalnaker's Thesis
For all Pr that model rational credence, and for all A, B such that A, B do not involve conditionals or modals:

$$Pr(A \to B) = Pr(B|A) \quad \text{(if } Pr(A) > 0\text{)}$$

[5] This might still be compatible with the claim that Modus Ponens is a valid argument form in some other sense. For example, it might be that Modus Ponens preserves support, even if it is not probabilistically valid.

Chapter 2

Triviality, 1/2: closure-based triviality proofs

How viable are Stalnaker's Thesis and Restricted Stalnaker Thesis?

> **Stalnaker's Thesis**
> For all Pr that model rational credence, and for all A, B:
>
> $$Pr(A \to B) = Pr(B|A) \quad \text{(if } Pr(A) > 0\text{)}$$
>
> **Restricted Stalnaker's Thesis**
> For all Pr that model rational credence, and for all A, B such that A, B do not involve conditionals or modals:
>
> $$Pr(A \to B) = Pr(B|A) \quad \text{(if } Pr(A) > 0\text{)}$$

Almost as soon as Stalnaker proposed it, various triviality results were proved, which seemed to show that it cannot hold (at least in full generality). What do these results really show? And how much of the thesis can non-trivially hold, in light of them?

Notice that the thesis generalizes in two dimensions:

- For any sentences A, B: every Pr is such that $Pr(A \to B) = Pr(B|A)$, if $Pr(A) > 0$.
- For any Pr: every pair of sentences A, B is such that $Pr(A \to B) = Pr(B|A)$, if $Pr(A) > 0$.

There are triviality results for both dimensions of generality. Thus, only in certain trivial cases can Stalnaker's Thesis hold fully generally across probability functions, and, at a single probability function, for every conditional sentence. Call triviality results of the first kind **closure-based proofs** and triviality results of the second kind **single probability proofs**. We will review both kinds in Chapters 2 and 3.

One thing you might have wanted was for Stalnaker's Thesis to hold, if at all, across the class of probability functions related by conditionalization. This seems plausible if you think that learning proceeds via conditionalization. Then, this would mean that your credence in a conditional should continue to equal your corresponding conditional credence even after you learn something new. Take the proposition expressed by $A \to B$, and suppose that given the probability distribution defined on your current evidence, Pr, is such that $Pr(A \to B) = Pr(B|A)$. Then, suppose you were to learn that X and thus update your credences via conditionalization:

- $Pr^{|X}$ is your probability distribution as updated by conditionalization; it is defined in terms of your original probability distribution as follows:
  For any Y, $Pr^{|X}(Y) = Pr(Y|X)$.

Then, we have:

- **Generalizing across conditionalization**: For any A, B and Pr:
  if $Pr(A \to B) = Pr(B|A)$, then $Pr^{|X}(A \to B) = Pr^{|X}(B|A)$, as long as $Pr^{|X}(A) > 0$.

Notice that this is a weaker claim than the fully general Stalnaker's Thesis as stated above. This claim is only that if you start with a thesis-conforming Pr, then you will end up with a thesis-conforming Pr even after you update via conditionalization.

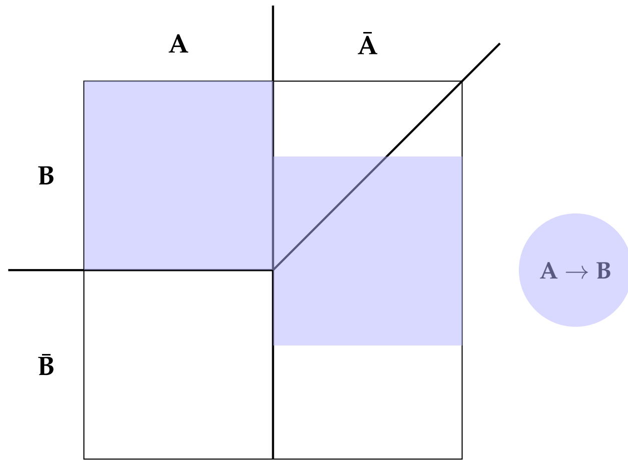
Unfortunately, not even **Generalizing across conditionalization** is true, except in trivial cases. This is the result established by Lewis's first triviality proof. Before we get to Lewis's proof, let's consider a warm up proof that will help us see why it works.

2.1 A warm up to Lewis

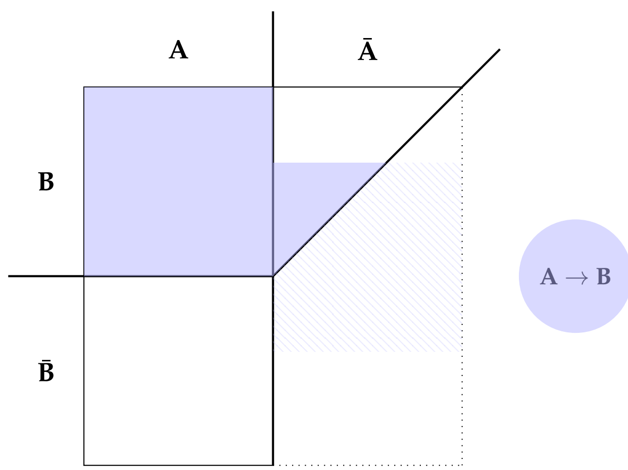
Recall from above that Probability Centering and Independence were sufficient to yield THE THESIS:

- PROBABILITY CENTERING: $Pr(A \to B \wedge A) = Pr(A \wedge B)$
- INDEPENDENCE: $Pr(A \to B|A) = Pr(A \to B)$

Graphically, this means that the conditional can be located in epistemic space roughly as follows:



The crucial properties are that the conditional is true throughout the **AB** region and nowhere within $\mathbf{A\bar{B}}$, and has a probability at the $\mathbf{\bar{A}}$-region that is equal to $Pr(B|A)$. Given this graphical representation, we can now easily see how to manipulate the probability distribution in ways that will not affect $Pr(B|A)$, but will affect $Pr(A \to B)$. One way is to conditionalize Pr on $\mathbf{A} \cup \mathbf{B}$:



Now, we have $Pr'(B|A) = Pr(B|A)$. But $Pr'(A \to B) < Pr(A \to B)$. So relative to Pr', we have a violation of **Generalizing across conditionalization**. What Lewis does with his proof is show that the only way to ensure that these kinds of failures of **Generalizing across conditionalization** won't happen (in certain non-trivial cases) is if $Pr(B|A) = Pr(B)$.

2.2 Lewis's first triviality result

Lewis proves the following triviality result:

LEWIS TRIVIALITY
If **Generalizing across conditonalization** holds, then for any sentences A, B, Pr such that $Pr(AB) > 0$ and $Pr(A\bar{B}) > 0$, $Pr(B|A) = Pr(B)$.

Proof.

Suppose $Pr(AB) > 0$, $Pr(A\bar{B}) > 0$ and $Pr(A \to B) = Pr(B|A)$.

By the law of total probability, we have:

(a) $Pr(A \to B) = Pr(A \to B|B) \cdot Pr(B) + Pr(A \to B|\bar{B}) \cdot Pr(\bar{B})$

Next, we have two lemmas (to be proved below):

(L1) $Pr(A \to B|B) = 1$
(L2) $Pr(A \to B|\bar{B}) = 0$

Substituting into (a) yields:

(b) $Pr(A \to B) = 1 \cdot Pr(B) + 0 \cdot Pr(\bar{B}) = Pr(B)$

And then by the fact that $Pr(A \to B) = Pr(B|A)$, it follows that:

(c) $Pr(B|A) = Pr(B)$

Proof of (L1). (The proof of (L2) is analogous)

For any Pr, A, B, X, Y (assume $Pr(AB) > 0$ throughout):

1. $Pr^{|X}(Y) = Pr(Y|X)$ — Definition: conditionalized probability function
2. $Pr^{|X}(A \to B) = Pr^{|X}(B|A)$ — THE THESIS
3. $Pr^{|B}(A \to B) = Pr(A \to B|B)$ — From 1
4. $Pr^{|B}(A \to B) = Pr^{|B}(B|A)$ — From 2
5. $Pr(A \to B|B) = Pr^{|B}(B|A)$ — From 3,4
6. $Pr^{|B}(B|A) = Pr(B|AB) = 1$ — See †
7. $Pr(A \to B|B) = 1$ — From 5,6

† Proof follows:

$$\mathrm{Pr}^{|B}(B|A) = \frac{Pr^{|B}(BA)}{Pr^{|B}(A)} = \frac{Pr(BA|B)}{Pr(A|B)} = \frac{\frac{Pr(AB)}{Pr(B)}}{\frac{Pr(AB)}{Pr(B)}} = \frac{Pr(AB)}{Pr(B)} \cdot \frac{Pr(B)}{Pr(AB)} =$$

$$\frac{Pr(AB)}{Pr(AB)} = Pr(B|AB) = 1$$

There are various reactions one may have to the proof:

- Deny step (a): this is to deny that *Pr* is a probability function, or else does not apply to conditionals.
  - This was Adams (1975)'s reaction, as well as Edgington (1995); Bennett (2003) following him in this tradition.
- Deny step (2): one might deny that Stalnaker's Thesis holds across probability functions closed under conditionalization.
  - Then a question arises: under what class of probability functions does Stalnaker's Thesis hold? Hájek's **triviality pursuit** result, discussed below, puts additional constraints on the class of functions across which the thesis can hold.
- Accept the conclusion: Stalnaker's Thesis holds only in trivial cases (material conditional).
  - This was Lewis (1976)'s reaction. He endorsed the material conditional analysis, according to which $Pr(A \to B) = Pr(B|A)$ only if $Pr(B|A) = 1$.

As we are going to see, merely denying Stalnaker's Thesis is not going to be sufficient though. Results that are analogous to Lewis's can be derived without any appeal to the Thesis.

2.3 Hájek's generalization

Lewis showed that Stalnaker's Thesis will fail for some conditionals and probability functions in any class of probability functions closed under conditionalization (assuming that the class contains some functions such that $Pr(AB) > 0$ and $Pr(A\bar{B}) > 0$). One question that arises at this stage is how general Lewis's proof is. For instance, can we get a similar triviality result for other classes of probability functions, those closed under a different operation that is not conditionalization? Lewis (1986) proved that a similar result hold for Jeffrey conditionalization. We won't review Lewis's second result here, but instead consider an even more general result, due to Hájek (2011b). Hájek's proof generalizes Lewis's result for a broad class of revision rules. Define two properties of revision rules:

- Let Pr_{*X} be the probability distribution *Pr* revised by *X* and rule $*$.
- A revision rule $*$ is **bold** iff for any *E* such that $Pr(E) > 0$, $Pr_{*E}(E) = 1$.
  - Bold rules revise into certainty. Conditionalization is one way to do this, but there are many other such rules. Here is one:

Extremism
If $Pr(X) > 0$, then $Pr_X(Y) = 0$, unless $X \models Y$, in which case $Pr_X(Y) = 1$.

- A revision rule $*$ is **moderate** iff for any A, E such that $A \models E$: if $Pr(A) > 0$, then $Pr_{*E}(A) > 0$.
  - Moderate rules require that revision preserves non-zero probability in anything that implies the proposition you are revising on. The extremist rule above is not moderate in this sense, because we could have $Pr(A) > 0$ and $A \models E$, but $Pr_E(A) = 0$ (as long as $E \not\models A$).
  - It seems plausible than any rational revision rule would be moderate. Suppose you previously thought it possible that it's raining hard, and then you learned that it's raining (and nothing else). It would be very strange if you could rationally now think it's not possible that it's raining hard.
- Say that Pr is non-trivial iff there are at least two sentences A and B such that $Pr(A) < 1, Pr(A \wedge B) > 0$ and $Pr(A \wedge \neg B) > 0$.

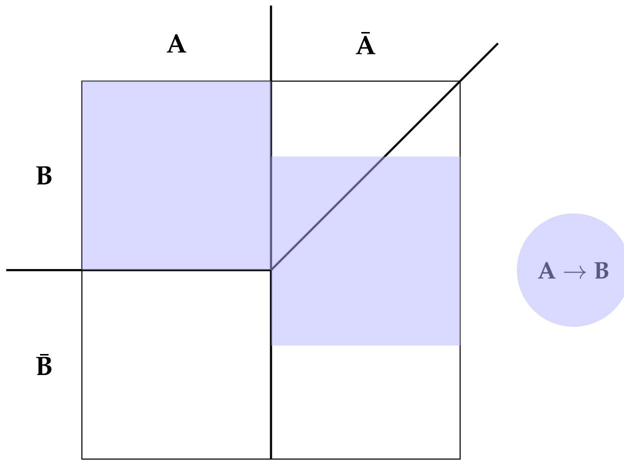
Hájek's proof establishes:

Hájek Triviality. If a class of probability functions is closed under a bold and moderate rule, then THE THESIS fails for some conditional in that class, unless it consists entirely of trivial probability functions.

I will skip one half of the proof, since I will assume Strong Centering:

STRONG CENTERING (SC)
$\models A \supset [(A \to B) \equiv B]$

Given (SC), it follows that $A \wedge B \models A \to B$, in which case there are no $A \wedge B$-worlds at which $A \to B$ is false. Conversely, $(A \wedge B) \wedge \neg(A \to B)$ is empty. The basic idea of the proof is straightforward. Start with a probability function that is non-trivial: thus, there are two sentences A and B such that $Pr(A) < 1, Pr(A \wedge B) > 0$ and $Pr(A \wedge \neg B) > 0$. Given (SC), the conditional $A \to B$ is true throughout $A \wedge B$ and false throughout $A \wedge \neg B$ and true somewhere in $\neg A$:



Assume that $Pr(A \to B) = Pr(B|A)$ (this is inessential to the proof, but assume it anyway). What we are going to do is revise Pr by $\neg(A \wedge B)$ (call this X) using a moderate and bold revision rule. As a result, it will be the case that $Pr_X(B|A) = 0$ but $Pr_X(A \to B) > 0$. Here is why. Since the revision is bold,

$$Pr_X(A \wedge B) = 0$$

Since it is moderate and $Pr(A \wedge \neg B) > 0$ and $A \wedge \neg B$ entails X,

$$Pr_X(A \wedge \neg B) > 0$$

and so $Pr_X(A) > 0$, and so

$$Pr_X(B|A) = 0$$

Next, since the revision rule is moderate, and $Pr((A \to B) \wedge X) > 0$ and $(A \to B) \wedge X$ entails X,

$$Pr_X((A \to B) \wedge X) > 0$$

But this implies that

$$Pr_X(A \to B) > 0$$

Thus, we have shown that there is a conditional $A \to B$ and probability function Pr such that we can construct a moderate and boldly revised Pr' such that THE THESIS fails for Pr' and $A \to B$. So, THE THESIS cannot hold across the class of probability functions closed under moderate and bold revision rules.

2.4 Bradley-style triviality proofs

Richard Bradley has contributed a number of triviality proofs. We present some of them below, together with some related results. Let us point out a general feature of Bradley-style proofs: differently from Lewis's and Hajek's proofs, they don't rely on Stalnaker's Thesis. Rather, they assume weaker constraints on credences in conditionals; these constraints are sufficient to derive implausible results. The moral is that giving up on Stalnaker's Thesis is not a sufficient response to triviality. On the contrary, these results suggest that there is a broader problem with the interplay of modal and probabilistic notions.

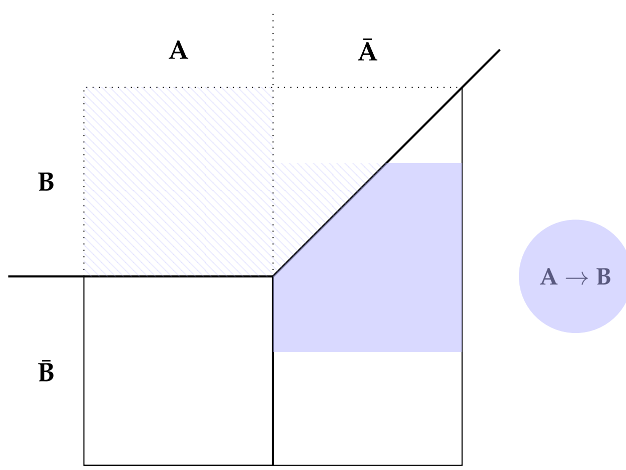
2.4.1 Bradley's preservation result

Bradley (2000) proves that, for any non-trivial language $\mathcal{L}$ containing a conditional operator (this is a language such that it contains sentences A, B and $A \to B$ and also $A \not\models B$ and $A \to B \not\models B$), the Preservation condition must fail for some $A, B \in \mathcal{L}$ and Pr.

- **Preservation condition**: If $Pr(A) > 0$ and $Pr(B) = 0$, then $Pr(A \to B) = 0$.

Proof. Let $\mathcal{L}$ be non-trivial. Then we have $A, B, A \to B \in \mathcal{L}$ and also $A \not\models B$ and $A \to B \not\models B$. Since $A \not\models B$, there is a Pr such that $Pr(A) > 0$ and $Pr(B) = 0$. And since $A \to B \not\models B$, there is a Pr such that $Pr(A \to B) > 0$ and $Pr(B) = 0$. Thus, it follows that there is a Pr such that $Pr(A) > 0$, $Pr(B) = 0$, and $Pr(A \to B) > 0$.

We can see graphically why this result holds by consulting the following diagram. The reason $A \to B$ has some positive probability is that $A \to B \not\models B$.



2.4.2 Bradley's recasting of Lewis

Bradley (2007) observes that, in Lewis's original result, the appeal to Stalnaker's Thesis only serves to derive the following two principles:

(L1) $Pr(A \to B|B) = 1$

(L2) $Pr(A \to B|\bar{B}) = 0$

Bradley then suggests that, rather than assuming Stalnaker's Thesis, we might just vindicate (L1) and (L2) via a different route. He suggest endorsing the following principle:

> *Cond-cert*. For any Pr modeling rational credence, then, if $Pr(A) > 0$:
>
> (a) If $Pr(C) = 1$, then $Pr(A \to C) = 1$
> (b) If $Pr(C) = 0$, then $Pr(A \to C) = 0$

Notice that there is intuitive motivation for *Cond-cert*, at least as long as we interpret '$\to$' as the indicative conditional. Consider the following examples:

(25) a. Frida is at the party.
b. If Maria is at the party, Frida is at the party.

Suppose that you regard (25a) as certain. Then it seems obvious that you should think (25b) is certain too. If you're certain that Frida is at the party, then, no matter what, you should also be certain that, if Maria (or for that matter anyone else) is at the party, then Frida is at the party. Similarly, *mutatis mutandis*, for the case in which (25a) gets probability zero.

Starting from *Cond-cert*, it's easy to prove that the probability of a conditional must equal the probability of the consequent.

1. $Pr(A \to C) =$
2. $Pr(A \to C \wedge C) + Pr(A \to C \wedge \bar{C}) =$ (total probability)
3. $Pr(A > C \mid C) \times Pr(C) + Pr(A \to C \mid \bar{C}) \times Pr(\bar{C})$ (def of $Pr(\cdot \mid \cdot)$)
4. $1 \times Pr(C) + 0 \times Pr(\bar{C}) =$ (via (a) and (b))
5. $Pr(C)$

What we have derived is the same result as Lewis, but without starting from the Thesis. This strategy can be replicated in a number of ways. In what follows, we explore a number of results that don't rely on the Thesis, and then draw some theoretical morals.

2.4.3 Triviality from Or-to-if: twisting Milne

A similar result can be obtained by modifying a result originally proven by Milne (2003). Milne presents a triviality proof to the effect that, on the assumption that $Pr(A \wedge C) > 0$, the probability of $A \to C$ is identical to the probability of the material conditional $A \supset C$.

Milne proves an auxiliary premise first. Assume that $Pr(A \wedge C) > 0$; we have:

1. $Pr(A \to C|A \supset C) =$
2. $Pr^{A \supset C}(A \to C) =$ (def of conditionalization)
3. $Pr^{A \supset C}(C|A) =$ (the Thesis)

4. $Pr(C|A \wedge (A \supset C)) =$ (via the proof of † in §2.2) =
5. $Pr(C|A \wedge C)) = 1$

Via a similar proof, we can establish:[1]

$$Pr(A \to C|\neg(A \supset C)) = Pr(A \to C|A \wedge \neg C)) = 0$$

At this point, we can run the usual proof that relies on total probability and the ratio formula:

1. $Pr(A \to C) =$
2. $Pr(A \to C \wedge (A \supset C)) + Pr(A \to C \wedge \neg(A \supset C)) =$
3. $Pr(A \to C|(A \supset C)) \times Pr(A \supset C) + Pr(A \to C|\neg(A \supset C)) \times Pr(\neg(A \supset C)) =$ (def of $Pr(\cdot \mid \cdot)$)
4. $1 \times Pr(A \supset C) + 0 \times Pr(\neg(A \supset C)) = Pr(A \supset C)$

Milne also show that, from the equivalence between $Pr(A \to C)$ and $Pr(A \supset C)$, we can derive all of Lewis's standard triviality results. So again we have triviality.

Here we want to point out that we can recast Milne's proof Bradley-style. We start from the following principle, which works as a counterpart of *Cond-cert*:

> *Prob-Or-to-If.* For any Pr modeling rational credence, then, if $Pr(A \wedge C) > 0$:
>
> (a) If $Pr(A \supset C) = 1$, then $Pr(A \to C) = 1$
> (b) If $Pr(A \supset C) = 0$, then $Pr(A \to C) = 0$

From here, we run a Bradley-style proof exactly analogous to the one above:

1. $Pr(A \to C) =$
2. $Pr(A \to C \wedge (A \supset C) + Pr(A \to C \wedge \neg(A \supset C)) =$ (total probability)
3. $Pr(A > C \mid A \supset C) \times Pr(A \supset C) + Pr(A \to C \mid \neg(A \supset C)) \times Pr(\neg(A \supset C))$ (def of $Pr(\cdot \mid \cdot)$)
4. $1 \times Pr(A \supset C) + 0 \times Pr(\neg(A \supset C)) =$ (via (a) and (b))
5. $Pr(A \supset C)$

Now, the reason why this proof is interesting is that both the sub-principles of *Prob-Or-to-If* follow from widely accepted principles about conditionals. Sub-principle (b) is a credal version of Modus Ponens. Sub-principle (a) is a probabilistic counterpart of well-entrenched principle about the acceptance of conditionals, namely Or-to-If (cf. Stalnaker 1975).

> **Or-to-If.** Any information stat that accepts $A \supset C$ also accepts $A \to C$.

[1] Milne does not rely on this premise to derive his conclusion. So far as I can see, his choice makes no difference at all.

While denying Stalnaker's Thesis might seem an option, it seems much less plausible to deny *Prob-Or-to-If*. This puts pressure on the other premises going into the proof, in particular closure and the ratio formula—in addition to the very idea that probability applies to conditional and modal statements in the first place.

2.5 General morals: triviality and informational inferences

The Bradley-style results suggest a general moral. Looking back at the proofs in §2.4.2 and §2.4.3, it's easy to notice that the particular content of the sentences figuring in *Cond-cert* and *Prob-Or-to-If* was not used at all. As a result, it's easy to generalize the proofs. Take any sentences X, Y; suppose that we have the following constraints:

a. If $Pr(X) = 1$, then $Pr(Y) = 1$
b. If $Pr(X) = 0$, then $Pr(Y) = 0$

Then, by using proofs structurally analogous to the ones above (and relying on closure under conditionalization), we can show that $Pr(X) = Pr(Y)$.

In fact, we can obtain results that are equally problematic just by assuming one of the two conditions above. Suppose we assume:

a. If $Pr(X) = 1$, then $Pr(Y) = 1$

This is sufficient to prove that the probability of X is a lower bound on the probability of Y (i.e. $Pr(X) \leq Pr(Y)$). (Provided that $Pr(Y \mid X) > 0$, $Pr(Y \mid \bar{X}) > 0$.)

i. $Pr(Y) = Pr(Y \wedge X) + Pr(Y \wedge \bar{X}) =$ (total probability)
ii. $Pr(Y \mid X) \times Pr(X) + Pr(Y \mid \bar{X}) \times Pr(\bar{X})$ (def of $Pr(\cdot \mid \cdot)$)
iii. $1 \times Pr(X) + Pr(Y \mid \bar{X}) \times Pr(\bar{X}) \geq Pr(X)$ (via constraint (a))

This alone is problematic in a number of cases. For example, it's problematic when X is $A \supset C$ and Y is $A \rightarrow C$, given the (near-)consensus that indicative conditionals are strictly stronger than material conditionals.

These observations establish a connection with another corner of the literature on epistemic modality and conditionals, i.e. that concerning so-called reasonable inference, or informational consequence (see, among many: Stalnaker 1975, Veltman 1996, Yalcin 2007, Bledin 2015). It is generally acknowledged that some inferences that are not classically valid are good inferences, in the following sense: whenever a subject accepts the premises, they are warranted in accepting the conclusion. As Stalnaker 1975, Or-to-If is exactly one such inference:

Or-to-If. $\varphi \vee \psi \vDash \neg\psi \rightarrow \varphi$

Another reasonable inference is the inference from C to $A \rightarrow C$, which of course is at the basis of Lewis's first triviality result, and Bradley's reformulation of it:

Consequent-to-If. $\psi \vDash \neg\psi \rightarrow \varphi$

(We refer the reader to the papers cited above for a formalization of the notion of reasonable inference, and its counterparts—so-called informational consequence and test-to-test consequence—in the dynamic semantics literature.)

It is a matter of contention whether reasonable inferences should be treated pragmatically or semantically. Whichever way one goes on this question, what matters to us is that reasonable inferences can be used to generate a kind of triviality. For (at least, if assigning probability 1 is sufficient for assigning a sentence) one will be able to run the proof schema above whenever we have a reasonable inference from X to Y. This will generally produce results that are incompatible with intuitive assignments of probability to conditionals.

This observation throws a natural bridge between triviality for conditionals and triviality for other kinds of modal claims. Reasonable inferences concern not just conditionals, but all kinds of modalized claims. Hence, if we can use reasonable inferences as a blueprint to generate triviality results, we should have triviality results for other modalized statements. This is exactly what we find.

2.6 Generalizing beyond conditionals: Russell and Hawthorne

We showed that triviality proof generalize in one dimension: we can get triviality from probabilistic constraints other than Stalnaker's Thesis. Now let us show that conditionals are not the only kind of modal claim that is subject to triviality. On the contrary, the literature contains triviality results for all kinds of modalized statements with epistemic flavor.[2]

Let's start by reviewing a triviality result for epistemic possibility modals originally presented by Russell and Hawthorne 2016 (henceforth R&H), in the context of providing a battery of triviality results for epistemic modals of all strength. I then show how this result can be strengthened into a collapse result that shows that A and $\Diamond A$ have the same probability.

R&H start from the following intuitive principle:

> **Might.** For any probability function Pr that models rational credence, if $Pr(\Diamond A) > 0$, then:
>
> $$Pr(A \mid \Diamond A) > 0$$

They show that, starting from **Might** and from the assumption that the class of rational credence functions is closed under conditionalization, we can prove that $\neg A$ and $\Diamond A$ are incompatible. Suppose for *reductio* that there is a probability function Pr on which $\neg A$ and $\Diamond A$ are compatible. Take $Pr_{\neg A}$, i.e. the result of conditionalizing Pr on $\neg A$. Via **Might**, we have:

i. $Pr_{\neg A}(A \mid \Diamond A) > 0$

By the definition of conditional probability, this means:

[2]And beyond: see Santorio 2018 for a triviality result about counterfactuals (also reported in chapter 7), and Santorio and Williams 2018 for a triviality result about the determinacy operator.

ii. $Pr(A \mid \Diamond A \wedge \neg A) > 0$

But via the probability calculus, we also have:

iii. $Pr(A \mid \neg A) = 0$

Lines (ii) and (iii) are inconsistent. So we conclude that there is no probability function on which $\neg A$ and $\Diamond A$ are compatible after all.

To be sure, the literature on epistemic modality includes notion of consequence on which $\Diamond A$ and $\neg A$ are treated as incompatible (see e.g. informational notions of consequence in the style of Yalcin 2007 or Bledin 2015). But it seems unacceptable that $\neg A$ and $\Diamond A$ should be *probabilistically* incompatible, i.e. that their conjunction should have probability zero. To better see this, let us notice that, when supplemented with an uncontroversial assumption, R&H's result leads to the absurd conclusion that A and $\Diamond A$ are probabilistically equivalent, i.e. they always get the same probability.

This is the uncontroversial assumption:

Box. For any probability function Pr that models rational credence:

$$\text{if } Pr(\Box A) = 1, \text{ then: } Pr(A) = 1$$

Box is the principle that assigning credence 1 to *must* A requires assigning credence 1 to A. This seems uncontroversial.[3] Now, we reason as follows. Take an arbitrary, rational probability function Pr. Via total probability, we have:

i. $Pr(\neg A) = Pr(\neg A \wedge \Diamond A) + Pr(\neg A \wedge \neg \Diamond A)$

Given the result of R&H's proof, we know that the first term goes to zero. Using the definition of conditional probability, we rearrange the second term as follows:

ii. $Pr(\neg A) = Pr(\neg A \mid \neg \Diamond A) \times Pr(\neg \Diamond A)$

Via the Duality of $\Box$ and $\Diamond$, this can be rewritten as:

iii. $Pr(\neg A) = Pr(\neg A \mid \Box \neg A) \times Pr(\neg \Diamond A)$

Now, since $Pr(\cdot \mid \Box \neg A)$ is a rational probability function (via closure under conditionalization), and since obviously $Pr(\Box \neg A \mid \Box \neg A) = 1$, via **Box** we know that $Pr(\neg A \mid \Box \neg A) = 1$. As a result, (iii) simplifies to:

iv. $Pr(\neg A) = Pr(\neg \Diamond A)$

But (iv) is disastrous, since it immediately leads to collapse. Via the probability calculus, we have:

[3] Notice that the principle doesn't amount to taking epistemic necessity to be veridical, which might indeed be controversial. It just requires that, if we are certain of *must* A, we are also certain of A.

v. $Pr(A) + Pr(\neg A) = 1$

From here, via (iv):

vi. $Pr(A) + Pr(\neg\Diamond A) = 1$

Via the probability calculus again:

vii. $Pr(A) + 1 - Pr(\Diamond A) = 1$

From here, rearranging:

viii. $Pr(A) = Pr(\Diamond A)$

This, of course, is absurd. (Just think of what probability you would assign to *It's raining* and *It might be raining* in cases in which you're not sure whether it's raining or not.)

The style of proof that we just gave generalizes to other modals. For an example, let us quickly rehearse R&H's proof about *Probably A*. The starting principle here is the following:

> **Probably.** For any probability function *Pr* that models rational credence, if $Pr(Probably(A)) > 0$, then:
>
> $$Pr(A \mid Probably(A)) > 1/2$$

As above, we start by supposing that, for some rational *Pr*, $\neg A$ and $Probably(A)$ are compatible (and hence have positive probability). As a result, by conditionalizing *Pr* on $\neg A$, $Probably(A)$ still gets positive probability:

i. $Pr_{\neg A}(Probably(A)) = Pr(Probably(A)|\neg A) > 0$

Now, we apply **Probably** to $Pr_{\neg A}$:

ii. $Pr_{\neg A}(A \mid Probably(A)) > 1/2$

But we also have:

iii. $Pr_{\neg A}(A) = Pr(A \mid \neg A) = 0$

Hence

iv. $Pr_{\neg A}(A \mid Probably(A)) = 0$

Contradiction between lines (ii) and (iv). So $Probably(A)$ and $\neg A$ must be incompatible after all.

Chapter 3

Triviality, 2/2: single probability triviality proofs

Closure-based triviality proofs undermine the generality of Stalnaker's Thesis for a single conditional across various probability functions. By contrast, single probability triviality proofs undermine the generality of Stalnaker's Thesis across a range of conditionals holding fixed a unique probability function. In this chapter, we explore some of the classic single probability triviality results of the last forty years, and the upshots for thinking about the probabilities of conditionals.

3.1 Fitelson 2015

Fitelson (2015) shows how to strengthen Lewis's proof so that it does not rely on any assumptions about probability functions related by conditionalization, but does rely on a probabilistic import-export principle:

> **Probabilistic Import-Export**: $Pr(A \to (B \to C)) = Pr((A \wedge B) \to C)$

Fitelson establishes that if **Probabilistic Import-Export** is valid, then:

> Import-Export Triviality: For every pair of sentences A, B and probability function Pr such that $Pr(AB) > 0$, $Pr(A\bar{B}) > 0$, and $Pr(B|A) \neq Pr(B)$, then one of the following instances of **Stalnaker's Thesis** must be false:
>
> (a) $Pr(A \to B) = Pr(B|A)$
>
> (b) i. $Pr((A \wedge B) \to B) = Pr(B|AB)$
>
> ii. $Pr((\bar{B} \wedge A) \to B) = Pr(B|\bar{B}A)$
>
> (c) i. $Pr(B \to (A \to B)) = Pr(A \to B|B)$
>
> ii. $Pr(\bar{B} \to (A \to B)) = Pr(A \to B|\bar{B})$

The proof is the same as Lewis's above, but offers an alternative proof of (L1)/(L2). We'll review the proof of (L1) (proof of (L2) is analogous, but uses (b-ii)/(c-ii) instead):

1. $Pr(B \to (A \to B)) = Pr((B \wedge A) \to B)$ **Probabilistic Import-Export**
2. $Pr((B \wedge A) \to B) = Pr(B|AB) = 1$ (b-i), $Pr(AB) > 0$
3. $Pr(B \to (A \to B)) = 1$ From 1,2
4. $Pr(B \to (A \to B)) = Pr(A \to B|B)$ (c-i)
5. $Pr(A \to B|B) = 1$ From 3,4

Here, the culprit is seems to be **Probabilistic Import-Export**. But the schema does seem valid . . .

(26) If the die landed on an prime, then if it landed on an even, it landed on 2.

(27) If the die landed on an even prime, it landed on 2.

(28) If Bob brought his umbrella, then if it isn't raining, Bob brought his umbrella.

(29) If Bob brought his umbrella and it isn't raining, Bob brought his umbrella.

These pairs seem equally likely. But what about? (from Kaufmann (2005); Fitelson (2016))

> Suppose that the probability that a given match ignites if struck is low, and consider a situation in which it is very likely that the match is not struck but instead is tossed into a camp fire, where it ignites without being struck. Now, consider the following two indicative conditionals.
>
> (a) If the match will ignite, then it will ignite if struck.
>
> (b) If the match is struck and it will ignite, then it will ignite.

See Khoo and Mandelkern (forthcoming) for further discussion about these issues.

3.2 Stalnaker 1976

Stalnaker (1976) contains a triviality proof for conditionals validating a restricted version of the thesis that the probability of a conditional equals the conditional probability of its consequent given its antecedent. The proof relies on four principles, and a non-triviality assumption:

Principles

Weak Stalnaker's Thesis (WST)
There exists a Pr such that for any A, B: $Pr(A \to B) = Pr(B|A)$.

Strong Centering (SC)
$\models A \supset [(A \to B) \equiv B]$

CONDITIONAL NON-CONTRADICTION (CNC)
$\models \neg[(A \to B) \wedge (A \to \neg B)]$

WELL-ORDER (CSO)
$A \to B, B \to A, A \to C \models B \to C$

Non-triviality assumption

- Say that Pr is **non-trivial** iff there are two propositions A, B such that $Pr(A) > 0, Pr(\bar{A}) > 0$ and $0 < Pr(B|A) < 1$.

The proof establishes that WEAK STALNAKER'S THESIS is true only of trivial probability functions. The proof makes crucial use of left-nested conditionals. Where A and B be two propositions for which Pr is non-trivial, we have:

- $0 < Pr(A \to B) = Pr(B|A) < 1$
- $C = A \vee \neg(A \to B)$
- $D = A \wedge \neg B$

We state the full proof below, but before we get there, we review Edgington (1995)'s helpful simplification. Let Pr be defined over four classes of worlds as per the following truth table:

w	$Pr(\{w\})$	A	B	$A \wedge \neg B (= D)$	$A \to B$	$A \vee \neg(A \to B) (= C)$	$C \to D$
1	1/4	T	T	F	T	T	F
2	1/4	T	F	T	F	T	T
3	1/4	F		F	T	F	F
4	1/4	F		F	F	T	F

The reason $A \to B$ is true at w_1, w_3 and false at w_2, w_4 is to ensure that $Pr(A \to B) = Pr(B|A) = 1/2$, respecting Strong Centering.

- By Strong Centering $C \to D$ is true at w_2 and false at w_1, w_4. But what about w_3? There it has a false antecedent and false consequent. We reason as follows:

  1. $A \to B$ — From truth table
  2. $(A \vee \neg(A \to B)) \to (A \wedge \neg B)$ — Reductio assumption
  3. $(A \vee \neg(A \to B)) \to A$ — 2, $\wedge$-elim
  4. $(A \vee \neg(A \to B)) \to \neg B$ — 2, $\wedge$-elim
  5. $A \to A$ — Trivial
  6. $A \to (A \vee \neg(A \to B))$ — 5, $\vee$-intro
  7. $(A \vee \neg(A \to B)) \to B$ — 1, 3, 6, CSO

8. $\perp$ 4, 7, CNC

- Thus, we conclude that $C \to D$ is false at w_3.

But now we can see that $Pr(C \to D) \neq Pr(D|C)$. $Pr(D|C) = 1/3$, and $Pr(C \to D) = 1/4$.

3.2.1 The full proof

Stalnaker's proof shows that this will hold no matter what the Pr you pick, as long as that Pr satisfies WEAK STALNAKER'S THESIS.

Suppose (WST), (SC), (CNC), (CSO), and (for reductio) suppose Pr is some non-trivial probability function satisfying (WST). Let A and B be the two propositions for which it is non-trivial. Thus, we have:

1. $0 < Pr(A \to B) = Pr(B|A) < 1$

Next, let

- $C = A \vee \neg(A \to B)$
- $D = A \wedge \neg B$

By (WT), we have:

2. $Pr(C \to D) = Pr(D|C)$

Next, we have:

3. $Pr(\overline{C}) > 0$

 3a. $Pr(A \to B|A) = Pr(B|A)$ By (SC)

 3b. Thus, $Pr(A \to B|\overline{A}) = Pr(B|A)$ By (WST), non-triviality, Probability Theory

 3c. So, $Pr(A \to B|\overline{A}) > 0$ and $Pr(\overline{A}) > 0$ (1), non-triviality, (3b)

 3d. So, $Pr(\overline{A} \wedge A \to B) > 0$ By (3c)

 3e. So, $Pr(\overline{C}) > 0$ By (3c) and since $\overline{C} = \overline{A} \wedge A \to B$

By similiar reasoning, we have:

4. $Pr(C \to D|C) = Pr(D|C) = Pr(C \to D|\overline{C})$

We use this, plus the following two lemmas, to derive a contradiction:

(L1) $Pr(C \to D|\overline{C}) = 0$

(L2) $Pr(D|C) > 0$

Proof of (L1): $Pr(C \to D|\overline{C}) = 0$

To begin, notice that:

5. $\overline{C} \models A \to B$.

We then prove:

6. $A \to B \models \neg(C \to D)$

   6a. Suppose for reductio that $A \to B$ and $C \to D$.

   6b. Then $C \to A, C \to \overline{B}$ — Since $D \models A \wedge \overline{B}$

   6c. Finally, we have $A \to C$ — Since $A \models A \vee \neg(A \to B)$

   6d. Thus, we have $C \to A$, $A \to C$, $C \to \overline{B}$.

   6e. So, it follows that $A \to \overline{B}$. — By CSO

   6f. So, we have a contradiction. — By CNC

So, by transitivity of entailment:

7. $\overline{C} \models \neg(C \to D)$

Therefore, it follows immediately that:

8. $Pr(C \to D|\overline{C}) = 0$

Proof of (L2): $Pr(D|C) > 0$

Start with the fact that, since $Pr(B|A) < 1$,

8. $Pr(A \wedge \overline{B}) > 0$

Next, we establish:

9. $A \wedge \overline{B} \models C \wedge D$.

   9a. Follows trivially, given that $C = A \vee \neg(A \to B)$ and $D = A \wedge \overline{B}$.

Therefore, by (SC) and transitivity of entailment:

10. $A \wedge \overline{B} \models C \to D$.

And hence,

11. $Pr(C \to D|D) > 0$. From (8), (9), (10)

12. $Pr(D|C) > 0$. From (4), (11)

3.3 Denying CSO?

A natural response to Stalnaker's result is to deny (CSO). For all we've shown so far, denying this might allow one to validate WEAK STALNAKER'S THESIS in full generality. Various theorists have explored this strategy, including Bacon (2015); Schulz (2014b). Both Bacon and Schulz drop the idea that the semantics of indicative conditionals involve orderings over worlds. Instead, they propose (roughly) that $A \to C$ is true iff a randomly selected A-world is a C-world. We'll leave the notion of randomness as a primitive (Bacon suggests that the randomly selected A-world just is the world that describes how things are if A is true; see Bacon (2015): 151). Given this, CSO fails:

- It can be the case that the randomly selected A-world is a B-world (so, $A \to B$ is true), and that the randomly selected B-world is an A-world (so $B \to A$ is true), and they are different worlds. This means that it can also be true that the randomly selected A-world is a C-world (so $A \to C$ is true), even though the randomly selected B-world is not a C world (so $B \to C$ is false).[1]

However, a serious challenge to this view arises from the fact that a restricted version of CSO is probabilistically valid, given:

> **Weak Stalnaker's Thesis**
> For all Pr that model rational credence, and for all A, B that do not involve conditionals or modals:
>
> $$Pr(A \to B) = Pr(B|A); \text{ if } Pr(A) > 0$$

Therefore, a random variable semantics must not validate **Weak Stalnaker's Thesis**, since it does not classically validate (and hence probabilistically validate) CSO. This puts pressure on us to find another response to Stalnaker's triviality result.

Here is the proof:

[1]Notice that for this reason, the random variable semantics predicts several other inference patterns fail; some of these failures are more plausible than others:

- Antecedent strengthening: $A \to C \models (A \wedge B) \to C$
- Antecedent weakening: $A \to C, B \to C \models (A \vee B) \to C$
- Consequent agglomeration: $A \to B, A \to C \models A \to (B \wedge C)$

- Preliminaries:
  - Let $U(A)$ be the uncertainty of A, given the relevant probability function, defined as follows: $U(A) = 1 - Pr(A)$.
  - $A_1, \ldots, A_n, \therefore C$ is probabilistically valid iff $U(C) \leq U(A_1) + \cdots + U(A_n)$.
    - Intuitively, this holds because if $A \models B$, then there cannot be more probability in the A-region than in the B-region: for that to hold, there would have to be A-worlds that are not B-worlds, in which case $A \not\models B$.

We aim to establish that:

> **OV**: CSO is probabilistically valid; that is: for no Pr is it the case that $U(B \to C) > U(A \to B) + U(B \to A) + U(A \to C)$.

Given LIMITED STALNAKER'S THESIS applies to each of these conditionals, it follows from **OV** that:

(*) For no Pr is it the case that $U(C|B) > U(B|A) + U(A|B) + U(C|A)$.

and hence:

(**) For no Pr is it the case that $1 - Pr(C|B) > 3 - Pr(B|A) - Pr(A|B) - Pr(C|A)$

Simplifying with algebra, we have:

(***) For no Pr is it the case that: $Pr(C|B) + 2 < Pr(B|A) + Pr(A|B) + Pr(C|A)$

This is our target. To prove this, assume it is false for reductio.

(R) There is a Pr such that $Pr(C|B) + 2 < Pr(B|A) + Pr(A|B) + Pr(C|A)$

We begin with a lemma:

> **Lemma 1**: For any Pr: $Pr(B|A) + Pr(C|A) \leq 1 + Pr(C|B)$

Proof of **Lemma 1**. Let Pr be an arbitrary probability function.

$$Pr(B|A) + Pr(C|A) = \frac{Pr(AB)}{Pr(A)} + \frac{Pr(AC)}{Pr(A)} =$$

$$\frac{Pr(AB) + Pr(AC)}{Pr(A)} = \frac{Pr(B) \cdot Pr(A|B) + Pr(C) \cdot Pr(A|C)}{Pr(A)}$$

Now, assume the best case here, that $Pr(A|B) = Pr(A|C) = 1$. This yields:

$$\frac{Pr(B)\cdot Pr(A|B)+Pr(C)\cdot Pr(A|C)}{Pr(A)}\leq\frac{Pr(B)+Pr(C)}{Pr(A)}$$

Next, assume the best case here, that $Pr(B)+Pr(C)=1$. This yields:

$$\frac{Pr(B)+Pr(C)}{Pr(A)}\leq\frac{1+Pr(C|B)}{Pr(A)}$$

Finally, assume the best case here, that $Pr(A)=1$. This yields:

$$\frac{1+Pr(C|B)}{Pr(A)}\leq 1+Pr(C|B)$$

Hence, from top to bottom, we have:

$$Pr(B|A)+Pr(C|A)\leq 1+Pr(C|B)$$

And this just is **Lemma 1**.

Now, we establish the falsity of (R).

- Let Pr^* be an arbitrary function that satisfies (R), so we have:

  (R1) $Pr^*(C|B)+2<Pr^*(B|A)+Pr^*(A|B)+Pr^*(C|A)$

- Since $Pr^*(A|B)\leq 1$, it follows from (R1) that:

  (R2) $Pr^*(C|B)+1<Pr^*(B|A)+Pr^*(C|A)$

- But **Lemma 1** yields:

  (R3) $Pr^*(C|B)+1\geq Pr^*(B|A)+Pr^*(C|A)$ By **Lemma 1**

- And (R2) contradicts (R3).

Given this result, there is an affinity between WEAK STALNAKER'S THESIS and (CSO), which bodes ill for the fully general form of STALNAKER'S THESIS. It seems that the best we're going to get is a limited version that doesn't apply to certain left-nested conditionals.

3.4 The Wallflower result

Hájek (1989) proves that, for any *Pr* such that STALNAKER'S THESIS holds for any A, B must have an infinite range. This means that there must be at least countably infinite values in $[1, 0]$ that *Pr* maps some sentence to. We won't look at Hájek's fully general result here, but instead explore a simpler way of understanding the result, due to Hájek (2011a). There, he calls this the "wallflower" result because it establishes that there must be some wallflowers—conditional probability values for *Pr*—not paired with any unconditional probability value for *Pr*. To see why, consider the following very simple finite model:

- $W = \{w_1, w_2, w_3\}$
- *Pr* is defined over $\wp(W)$ such that $Pr(\{w_1\}) = Pr(\{w_2\}) = Pr(\{w_3\}) = 1/3$.
- Therefore, the probability of any proposition in $\wp(W)$ must then be a multiple of $1/3$.
- However, various conditional probabilities are not multiples of $1/3$.
  - Let **A** = $\{w_1, w_2\}$ and **B** = $\{w_1\}$. Then, $Pr(B|A) = \dfrac{Pr(\{w_1\})}{Pr(\{w_1, w_2\})} = \frac{1/3}{2/3} = 1/2$.

What should we make of this result? As we'll see, one way to resist it is to block the most straightforward connection between *Pr* and *P*. This is something we'll return to when we discuss the tenability results.

Chapter 4

Resisting Triviality

4.1 Strategy 1: nihilism

The nihilist strategy concludes that the offending principles, STALNAKER'S THESIS, or the PRESERVATION PRINCIPLE are not just invalid, but they are not even close to being valid. We'll consider two nihilist responses and then consider how a nihilist might adopt the insights of the restrictor theory to offer a plausible account of why these principles *seemed* valid in the first place.

4.1.1 Probability and assertability

Lewis's own response to his triviality proof was to endorse the material conditional theory, and thus concluded that Stalnaker's Thesis holds only in trivial cases where $Pr(C|A) = 1$ or $Pr(C|A) = 0$. But how, then, does he account for the intuitions behind Stalnaker's Thesis (and also the Bradley principles)? Lewis (along with Jackson (1979, 1987)), held that these intuitions track epistemic assertability rather than probability.

We spell out this strategy below. But before we do so, we want to point out that this assertability strategy is available to anyone, not just material conditional theorists: in fact, one of the first expressvists about conditionals (Ernest Adams) endorsed something like this strategy as well (see Adams (1965)).

Lewis offered a Gricean argument for why high probability of truth may not be enough for assertability:

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

> ditional reason to believe that the conditional is true, over and above his disbelief in the antecedent. (Lewis (1976): 306)

So what else is needed for a conditional to be assertable? Jackson, who further developed this response, suggested:

> **Assertability (Jackson)**
> $A \to B$ is assertable to a degree equal to $Pr(B|A)$, where $Pr(A) > 0$.

Think of degrees of assertability as falling in the unit interval, with degree 1 being perfectly assertable and degree 0 being perfectly unassertable, and degrees in between being more or less assertable.

Jackson's condition offers a potentially promising error theory of our judgments about the probabilities of conditionals. The idea is this: when we seem to think that the probability of $A \to B$ is equal to the probability of B given A, we are really judging its degree of assertability.

How does this avoid the triviality proofs? The theory claims that $Pr(A \to B) \neq Pr(B|A)$, but $D(A \to B) = Pr(B|A)$, where $D(A \to B)$ is the degree of assertability of the conditional. And we may suppose that degrees of assertability do not obey the law of total probability, so we block the attempted revenge proof at step 1.

How plausible is **Assertability (Jackson)**? Consider an example:

(30) If I toss this fair coin, it will land heads.

How assertable is this? Jackson says that it is 0.5 assertable. But Keith DeRose suggests otherwise:

> To the extent that I can just intuit the degree to which the conditional is assertable, I would give it a value much lower than 0.5. (Forced to assign a number, I would go for something like 0.06.) After all, it is a fair coin. So I have no idea which of its two sides it will land on if I toss it. I would have to say that I am in no position to assert either that it will land heads if I toss it, or that it will land tails if I toss it. And it does not seem a close call: neither conditional seems close to being half-way assertable. (DeRose 2010: 12)

Lottery cases provide even more striking counterexamples. Suppose I entered a fair 1000 ticket lottery.

(31) If the drawing was held, I lost.

One feels in this case a strong intuition that (31) is not assertable at all. Maybe it's a bit more assertable than the coin case, but it's nowhere near perfect assertability. Yet, since Pr(Lost | Drawing held) = .999, it is predicted to have a very high degree of assertability, and hence should seem very assertable.

If these intuitions are on the right track, this suggests that **Assertability (Jackson)** is wrong, and thus cannot account for our intuitions about the probabilities of conditionals. After all, the probability of (30) is intuitively 1/2, and the probability of (31) is intuitively .999.

The lesson is this: probability and assertability are different! We cannot account for our intuitions about probabilities of conditionals by appealing to assertability.

4.1.2 Expressivism

One common strategy of response to triviality endorses the claim that conditionals and epistemically modalized sentences don't express standard propositions, but rather have contents of other sorts. This response has been originally pursued in the literature on philosophical logic by Ernest Adams (see e.g. Adams 1996) and Dorothy Edgington (see e.g. Edgington 1995).

The basic idea is simple: if conditionals or modalized claims don't express propositions, then standard probability measures are simply not defined over conditionals or their contents. As a result, a basic assumption that we need to formulate any kind of triviality results is blocked.

This said, expressivist accounts need to offer *some* explanation of the probability judgments that ordinary subjects express about conditionals. It is at this point that expressivists split into two strands.

- **All-out nihilists** hold that conditionals and modalized sentences have no probability, and moreover no credence-type notion applies to them. Edgington is the most obvious defender of this view. (Though, in a different way, also Seth Yalcin may be counted in this group.)

- **Moderate nihilists** hold that conditionals and modalized sentences have no probability, but they are the object of some credence-type attitude, which is distinct from probability. This attitude can be understood in various terms. Ernest Adams and (in unpublished work) Nate Charlow are defenders of this view.

Worries for all-out nihilists. The clearest form of all-out nihilism is the view defended by Dorothy Edgington. Edgington claims that conditionals don't express propositions. Rather, they are linguistic devices for expressing the fact that the speaker assigns a high conditional probability to the consequent, given the antecedent.[1] As such, conditionals don't have probabilities themselves. Rather, they are just used to express those probabilities.

The major worry about Edgington's view is that it does not have the resources to account for the compositional properties of conditionals. Edgington simply doesn't assign anything like a semantic value to conditionals. Perhaps the view can be supplemented in various ways (possibly with the compositional resources of modern expressivism). But, as it is, it clearly fails to capture something crucial about conditionals in natural language.

[1] An analogy: roughly, normative expressivists like Gibbard hold that normative claims like *Murder is wrong* don't express propositions, but rather are just devices for expressing one's endorsement or rejection of a certain course of action.

A second worry is that hardwiring the connection to conditional probabilities in the meaning of conditionals might be too strong. On Edgington's view, it is simply a conceptual truth that $A \to C$ expresses that the speaker assigns a high conditional probability in the consequent, given the antecedent. Hence any counterexamples to the Thesis are, *prima facie* at least, direct counterexamples to the view. Both McGee-style examples (like the ones we discussed in chapter 1) and Kaufmann-style examples are potential worries for Edgington.

Worries for moderate nihilists. The main task for moderate nihilists consists in providing a plausible accounts of the credence-type notion that applies to conditionals. One option is to understand this notion in terms of assertability; this option is subject to the same objections discussed in §4.1.1. Other attempts will need to be examined on a case-by-case basis.

Let us just mention one general concern for this kind of nihilist. On the one hand, the relevant credence-type notion will presumably need to be related to probability in some way. For example, the expressivist should vindicate the centering entailments:

$$A \wedge C \vDash A \to C \vDash A \supset C$$

Hence, presumably, we should get that the credence in $A \to C$ is somehow bounded from above by the credence in $A \wedge C$ and from below by the credence in $A \supset C$.

At the same time, the relevant notion of credence should be distinct enough from ordinary probabilities to prevent us from reproducing triviality proofs. It is an interesting question whether one can find a credence-type notion that satisfies both these constraints.

4.1.3 Invoking the restrictor analysis

Recall Kratzer's formulation of the restrictor theory, as discussed in Chapter 1:

Restrictor Theory
$[\![A \to B]\!]^{w,f} = 1$ iff $[\![B]\!]^{w,f^{\mathbf{A}}} = 1$

Let's lift this theory into one that evaluates sentences relative to an information state i (a pair of a set of worlds and a measure on them):

Restrictor Theory*
$[\![A \to B]\!]^{w,i} = 1$ iff $[\![B]\!]^{w,i^{\mathbf{A}}} = 1$

What is $i^{\mathbf{A}}$? Following Yalcin (2010), we define it as follows:

$$i^{\mathbf{A}} = \langle s_i \cap \mathbf{A}, P_i^{|\mathbf{A}} \rangle$$

Now, let's give a semantics for *probably*, as follows:

$$[\![\textit{Probably } A]\!]^{w,i} = 1 \text{ iff } P_i(\mathbf{A}) > .5$$

With these two definitions in hand, we can now show that conditionals with probability operators in their consequents will be true or false depending on whether the corresponding instance of the THESIS is true or false. I.e., we have:

$[\![\textit{if A, probably B}]\!]^{w,i} = 1$ iff

$[\![\textit{probably B}]\!]^{w,i^{\mathbf{A}}} = 1$ iff

$P_{i^{\mathbf{A}}}(\mathbf{B}) > .5$ iff

$P_i^{|\mathbf{A}}(\mathbf{B}) > .5$ iff

$P_i(\mathbf{B}|\mathbf{A}) > .5$

(In fact, given that we can modify *probably* with various numerical determiners, we can have the following: the object language sentence *It is n likely that, if A, then* C is going to be true just in case $Pr(C \mid A) = n$.)

Crucially, this result holds regardless of any of the triviality results above, since the truth value of *if A, probably B* does not depend on the unconditional probability of any proposition, but rather entirely on the conditional probability of B given A.

Replicating triviality for Kratzer. Let us be clear about what the above shows. One large part of the motivation for the Thesis is the cogency of claims like the following:

(32) It it 50% likely that, if Maria tossed the coin, the coin came up tails.

Any theory of conditionals that fails to explain why speakers find 32 plausible misses an important empirical point. Now, Kratzer's restrictor theory can account for this. On a plausible semantics for *likely*, 32 will turn out to be true.

But this does not mean that Kratzer gets out of triviality results (as has been pointed out explicitly by Charlow 2016). The fact that 32 is true doesn't entail that, once we assign probabilities to object language sentences *qua* theorists, we get the right results for 33:

(33) If Maria tossed the coin, the coin came up tails.

In fact, the restrictor strategy does nothing at all to address classic triviality results. This means that, if we stick with Kratzer, as long as we validate one of the constraints on conditionals probability that lead to triviality (the Thesis, Preservation, *Cond-Cert*, *Prob-Or-to-If*), we will be able to run the proofs exactly like we ran them in previous chapter.

So the restrictor response amounts to a sophisticated error theory about the Thesis. It is able to explain away a large amount of intuitions about probabilities of conditionals. But (unless we block the proofs in some other way) it does not allow us to prevent the probabilities of the propositions expressed by conditionals from being trivial.

Even as an error theory, however, the restrictor theory seems to us to suffer from serious problems. Let us mention two of them.

Trouble from propositional anaphora. First, the strategy is implausible if we attend to the way that propositional anaphora works in natural language. If we focus on propositional anaphora like *that*, this may be plausible:

A: If the die landed on a prime, it landed on an odd.

B: That's likely.

After all, it's often unclear whether *that* targets a matrix clause or an embedded clause:

A: I think it's raining.

B: That's likely. ($\neq$ It's likely that you think it's raining)

However, if we focus on other kinds of propositional anaphora, the strategy becomes less plausible. Consider:

A: If the die landed on a prime, it landed on an odd.

B: What A said is likely.

Note first that the intuition seems the same in this case as in the prior one: B's claim is true just if it's likely that the die landed on a prime given that it landed on an odd. But in this case, A hasn't said that it landed on an odd, so B's anaphor "what A said" cannot refer to that proposition. Rather, the only plausible thing it could be referring to is the content of A's claim: that is, the content of the conditional "if the die landed on a prime, it landed on an odd." But if this is so, then it's hard to see how the restrictor theory strategy could work, for the strategy was designed precisely to account for our THESIS-friendly intuitions without predicting that they are judgments of the probability of the content of the conditional.

Trouble from compounds of conditionals. The second problem is that the strategy cannot handle compounds of conditionals. To see this, consider the following scenario:

> *Coins.* Martina is considering tossing two fair coins, A and B, in two independent tosses. You leave the room before you discover whether she tosses them or not.

Now, assess the probability of the following statements:

(34) Coin A landed heads, if it was tossed, and coin B landed tails, if it was tossed.

(35) Each of coin A and coin B landed heads, if it was tossed.

One natural judgment, which has been confirmed by several speakers, is that both (1) and (2) have probability $1/4$. But this judgment cannot be vindicated via the restrictor strategy. To see this, consider the object language sentences that we obtain if we prefix (34) and (35) with the object language locution *It is 1/4 likely that*.

(36) It's 1/4 likely that coin A landed heads, if it was tossed, and coin B landed tails, if it was tossed.

(37) It's 1/4 likely that each of coin A and coin B landed heads, if it was tossed.

Crucially, on a Kratzer-style analysis we cannot take the *if*-clauses in (36) and (37) to restrict the probability operator. The logical forms of (36) and (37) are, at the relevant level of abstraction:

(38) 1/4 LIKELY [[if Coin A was tossed, it landed tails]] $\wedge$ [[if Coin B was tossed, it landed tails]]

(39) 1/4 LIKELY [For every coin x. [if x was tossed] [x landed tails]]

In both cases, there is an item occurring between the probability operator and the *if*-clause or clauses (a conjunction in (36) and a universal determiner in (37)).[2] Hence the latter cannot work as restrictors of the former.

4.2 Strategy 2: context dependence (overview)

A second broad strategy to block a large class of triviality proofs is to invoke context dependence. This strategy is effective against closure-based triviality proofs, which exploit different probability functions. The leading idea is that information-sensitive expressions, which include epistemic modals and conditionals, express different propositions in different epistemic contexts. In particular, an information-sensitive sentence will shift its meaning across information states.

First, let's get clear about what the contextualist denies. Since they accept that conditionals express propositions that obey Probabilistic Centering and Independence, they deny (L1)/(L2):

(L1) $Pr(A \to B|B) = 1$

(L2) $Pr(A \to B|\bar{B}) = 0$

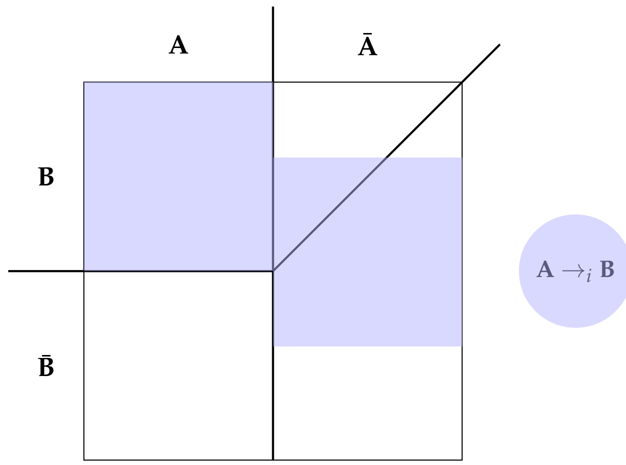
In so doing, the contextualist blocks Lewis's/Bradley's proofs, which both appeal to these principles.

Why do these fail? Focus on the following example:

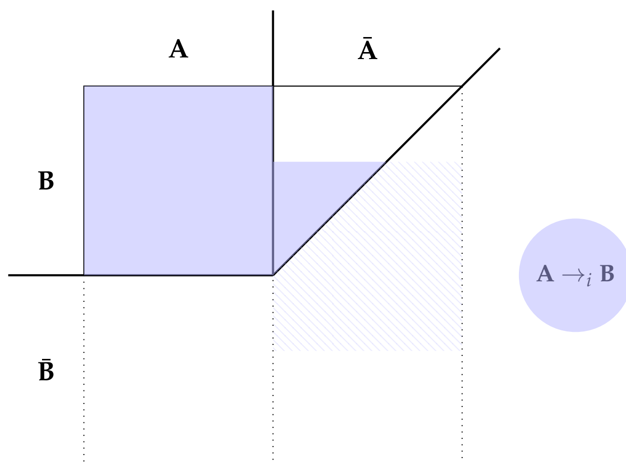
[2] Wolfgang Schwarz pointed out to one of us (P.S.) that a Kratzer-style picture would get the correct truth conditions if it allowed for the restrictors of both conditionals to *jointly* restrict the probability operator. This suggestion runs into two objections, both of which we take to be conclusive. First, to our knowledge there is no syntactic view that allows a conditional antecedent to outscope a conjunction and go restrict a higher modal. Second, the view predicts that a conjunction of conditionals with incompatible antecedents should have at least a prominent reading on which it is judged to have probability zero. But this seems incorrect. Consider:

(1) Maria is in her office, if the lights are on, and she is in her office, if the lights are not on.

We cannot see any reading of (i) on which it gets probability zero.



Here is the functional conditionalized on **B**:



Notice here, that $Pr_i(A \rightarrow_i B|B) \neq 1$. So, we deny (L1). But wait, didn't we get an argument for (L1) by Bradley's Cond-cert principle? Yes!

> *Cond-cert*. For any Pr modeling rational credence, then, if $Pr(A) > 0$:
>
> (a) If $Pr(C) = 1$, then $Pr(A \rightarrow C) = 1$
>
> (b) If $Pr(C) = 0$, then $Pr(A \rightarrow C) = 0$

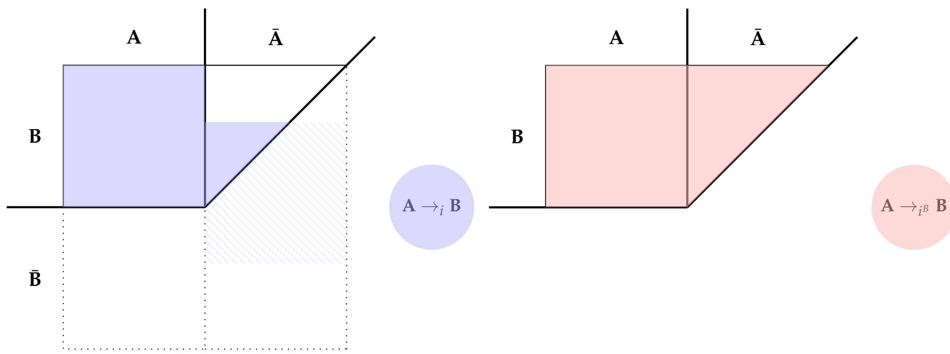
Thus, the contextualist denies Cond-cert. But, then, why did Cond-cert seem plausible in the first place? The contextualist strategy is to hold that, even though Cond-cert is false, a nearby, but weaker, principle is true.

- For any information state i, let $[\![A]\!]^i = \{w\colon [\![A]\!]^{i,w} = 1\}$
- Use '$\mathbf{A} \rightarrow_i \mathbf{B}$' as a shorthand for $[\![A \rightarrow B]\!]^i$.

The weaker principle is one which quantifies not over probability functions, but information states:

> *Local cond-cert.* For any i, if $P_i(\mathbf{A}) > 0$:
>
> (a) If $P_i(\mathbf{C}) = 1$, then $P_i(\mathbf{A} \to_i \mathbf{C}) = 1$
>
> (b) If $P_i(\mathbf{C}) = 0$, then $P_i(\mathbf{A} \to_i \mathbf{C}) = 0$

How do Cond-cert and Local cond-cert come apart? They come apart because $P_i^{|B}(A \to_i B) \neq P_{i^B}(A \to_{i^B} B)$. The key is that changes to i lead to changes in $[\![A \to B]\!]^i$. Compare:



$P_i(\mathbf{A} \to_i \mathbf{B} \mid \mathbf{B}) < 1$ $\qquad$ $P_{i^B}(\mathbf{A} \to_{i^B} \mathbf{B}) = 1$

Is there a semantics for conditionals that validates Local cond-cert? Yes, many do! Here is one:

> **Strict**
> $[\![A \to B]\!]^{i,w} = 1$ iff for all $w \in s_i \cap \mathbf{A}$: $[\![B]\!]^{i,w'} = 1$.

Suppose that $P_i(\mathbf{C}) = 1$. Then all of the worlds in s_i are **C**-worlds. So every **A**-world in s_i is a **C**-world. So $\mathbf{A} \to_i \mathbf{C}$ is true at every world in s_i. So $P_i(\mathbf{A} \to_i \mathbf{C}) = 1$. The same reasoning goes for the other case (from $P_i(\mathbf{C}) = 0$ to $P_i(\mathbf{A} \to_i \mathbf{C}) = 0$).

4.2.1 Troubles for contextualism

We consider two issues for contextualism. The first is that, while contextualism is not implausible in the epistemic domain, it might be less plausible for other domains. Given how easily we can generate triviality results, we might run into one that is not easily amenable to a contextualist treatment. Some examples include: triviality results for counterfactuals (see below) and determinacy operators.

The second problem facing contextualism concerns its prediction for the probabilities of conditionals across contexts with different information. Suppose that Maria tossed a fair die, and that neither you nor Frida have information about the outcome. Frida then says:

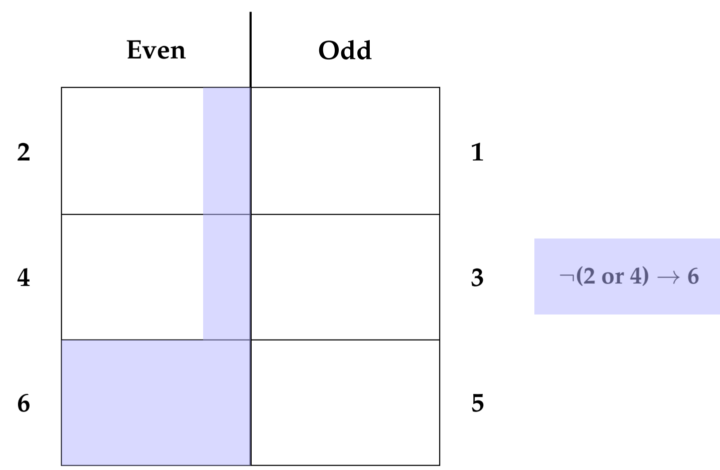
(40) If the die didn't land on two or four, it landed on six.

Intuitively, Pr_i(**not (two or four)**$\rightarrow_i$ **six**) = 1/4. But now suppose you learn (41):

(41) The die landed even.

What is the probability of what Frida said? One very natural answer here is 1—you should now be certain of what Frida said, given that you now know that the die landed even. However, this is not what contextualism predicts.

To see why, consider the following diagram:



Notice that, by Centering, the conditional ¬**(2 or 4)** → **6** is true if a six is rolled, and false if a 1, 3, or 5 is rolled. And if a 2 is rolled, it has probability 1/4 (and same if a 4 is rolled). This ensures that P_i(¬**(two or four)** $\rightarrow_i$ **six**) = 1/4, by total probability.[3]

Now, when we conditionalize on **even**, we eliminate the **Odd** region as follows, yielding the new function $P_i^{|\textbf{Even}}$:

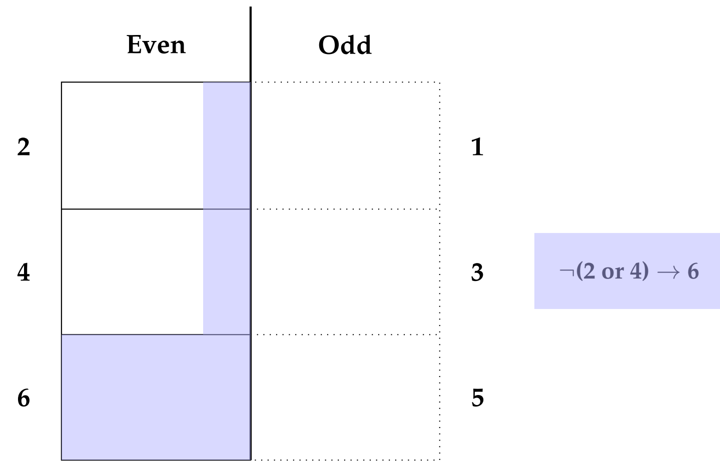
[3]Here is the calculation:

- P_i(¬ **(two or four)** $\rightarrow_i$ **six**) =

  P_i(¬ **(two or four)** $\rightarrow_i$ **six** | **even**) · P_i(**even**) + P_i(¬ **(two or four)** $\rightarrow_i$ **six** | **odd**) ·P_i(**odd**) =

  P_i(¬ **(two or four)** $\rightarrow_i$ **six** | **even**) $\cdot 1/2 + 0 = 1/4$

  So, P_i(¬ **(two or four)** $\rightarrow_i$ **six** | **even**) = 1/2



Crucially, notice that $P_i^{|\textbf{Even}}(\neg$ **(two or four)** $\rightarrow_i$ **six**) = 1/2. Thus, if the expression "What Frida said" refers to the proposition expressed by Frida's utterance of (40), then contextualism seems to predict the wrong judgments.

How bad a result this is for contextualism depends on the how plausible the responses are in defense of the view. We will not go into that issue here. For a possible defense of a related kind of contextualism from challenges involving propositional anaphora, see Khoo (2018).

4.3 Strategy 3: denying Ratio or closure

4.3.1 The options: the ratio formula and closure

The next option to resist triviality is to resist one of the assumptions that are involved in transitioning from one step to the other in triviality proofs. There are two assumptions that are common to all closure-based triviality proofs:

Ratio. $Pr(C \mid A) = \dfrac{Pr(A \wedge C)}{Pr(A)}$

Closure under conditionalization: For any X and Pr:
if $Pr(\cdot)$ models a rational credence distribution, then $Pr(\cdot|X)$ models a rational credence distribution.

Notice that Ratio and Closure are pretty much the only things we need to generate a Bradley-style result. For illustration, here is a minimal triviality proof that we can get exploiting just the first half of *Prob-Or-to-If*. Beyond Ratio and Closure, we are going to use the following:

Lower Bound. For any X, $Pr(A) \geq Pr(A \wedge X)$

Lower Bound is a basic principle saying that the probability of a conjunction is a lower bound on the probability of a conjunct. It is a weakening of the principle of total probability which was invoked by standard Bradley proofs.

1. $Pr(A \to C) \geq$
2. $Pr(A \to C \wedge (A \supset C) =$ (Lower Bound)
3. $Pr(A > C \mid A \supset C) \times Pr(A \supset C) =$ (Ratio)
4. $1 \times Pr(A \supset C) \times Pr(\neg(A \supset C)) =$ (*Prob-Or-to-If*, Closure)
5. $Pr(A \supset C)$

This suggests that, unless we go for context dependence or a nihilist option, we should really focus on one of Ratio and Closure, as opposed to e.g. total probability. Let's see what the prospects are for denying one of them.

4.3.2 Denying Ratio

Sometimes Ratio is presented merely as a definition of conditional probability. On this understanding, of course denying it makes little sense: if we deny it, we are merely deciding to change the subject and use a different notion of conditional probability. On the other hand, we can also interpret Ratio in a different way. We can take conditional probabilities to be partly defined by the rule of conditionalization.

> **Conditionalization.** $Pr_A(C) = Pr(C \mid A)$

On this understanding, conditional probabilities are individuated as those probabilities that are used to replace one's prior probabilities after updating on new evidence. If we understand conditional probabilities in this way, then Ratio is a substantial claim, and can be denied. What we're denying is that update proceeds via conditionalization.

This might seem to be a radical step. Or at least, it might seem radical to take this step for the purpose of rescuing our intuitions about probabilities of conditionals and modal claims. But there might be principled reasons to do so. Let us gesture towards some of them.

A number of classical logical relations validate the following principle ('$i \vDash \varphi$' stands for 'information state i' makes true sentence φ):

> **Persistence.** If $i \vDash \varphi$, then, if $i^+ \subseteq i$, $i^+ \vDash \varphi$.

Persistence simply says that, if an information state makes true a sentence φ, then any information state that is at least as strong as it validates φ. It has been claimed in the literature that epistemically modalized sentences are precisely a counterexample to this claim (see, among many, Veltman 1996). Suppose that i validates *might* A. $i \cap \mathbf{A}$ is a state that is more informed than i, and fails to validate it.

Richard Bradley provides a similar example for conditionals, pointing out that $\neg A \to B$ is validated by any information state that validates $A \vee B$ without validating any of the disjunct, but not by information states validating A.

> When I learn that the prize is in either urn A or urn B, I infer that if it is not in A then it must be in B. But when it is subsequently revealed that the prize is in A, I learn that the truth of the claim that it was

> in either A or B derived from that fact that it was in A and so my grounds for the inference that if it was not in A then it was in B is removed. None the less, [Persistence] requires me to retain it.

Now, a probabilistic analog of Persistence is obviously entailed by Conditionalization.

> **Probabilistic persistence.** If $Pr(A) = 1$, then, for any X: $Pr(A \mid X) = 1$

So, if we take seriously the idea that epistemically modalized claims invalidate Persistence, then we need to take equally seriously the idea that they invalidate Conditionalization. Of course, it is an open question how to formulate a replacement for conditionalization.

4.3.3 Denying Closure

The second step is denying closure. On a first pass, denying closure seems radical. It seems that, to deny it, we should deny that by observing A and conditionalizing on it, we are guaranteed to land on a rational belief state. This seems a straightforward violation of a classical Bayesian postulate.

The first thing to observe is that classical Bayesianism was never intended to characterize the notion of rationality that we're interested in now. Recall that we charaterized the Thesis as a rational constraint that (obviously) goes beyond classical subjective Bayesianism. While it might be very surprising that conditionalization fails, it does not strictly speaking refute classical Bayesian postulates.

Second, notice that failure of closure can be characterized in two subtly, but importantly different ways:

i. There is a proposition p such that: if p is an agent A's total evidence and A conditionalizes on p, A's resulting credal distribution is non-rational.

ii. There is a proposition p such that: p is never a rational agent's total evidence.

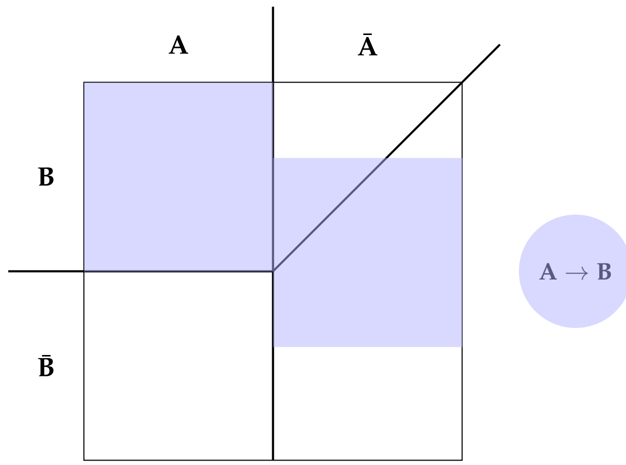
Option (ii), we suggest, is the most plausible way of understanding the denial of closure. According to this option, we have to impose restrictions on what an agent's total evidence is. This can be seen less a claim about rationality, and more as a claim about the space of meanings for sentences in our language. Somehow, some sentences cannot be learned in isolation; learning them invariably involves learning also some other sentences.

One simple illustration of this is given by A and *mustA*. Whenever A is learned, *mustA* is learned as well. One cannot learn A without learning *mustA*, despite the fact that the former does not entail the latter (at least not on any notion of entailment that works well in combination with probability).

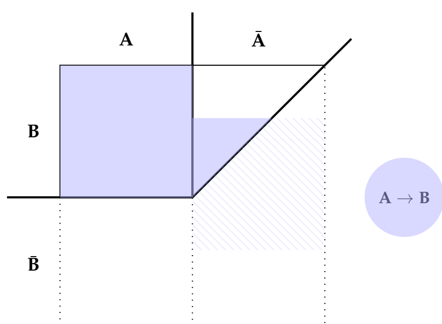
As for the denial of Ratio, this avenue also requires further justification and theorizing. But it seems a live option, especially in combination with a non-truth-conditional understanding of conditionals and epistemic modals.

4.3.4 Issues to think about

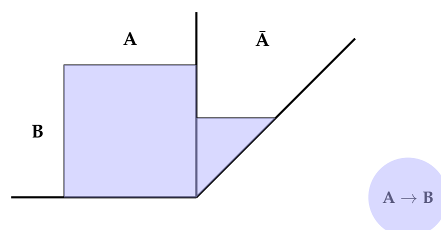
One way of thinking about this strategy is that, when you learn some factual proposition **A**, you thereby learn a host of conditional propositions: for each $\mathbf{X} \cap s_i \cap \mathbf{A}$, you learn $\mathbf{X} \to \mathbf{A}$. So, start out with our conditional in logical space:



Now, when we learn **B**, we don't conditionalize: that yields the left case, which violates Stalnaker's Thesis. Rather, we update by this other special learning process whereby we come to learn $\mathbf{A} \to \mathbf{B}$.



$P_i(\mathbf{A} \to \mathbf{B}|\mathbf{B}) < 1$



$P_{i^B}(\mathbf{A} \to \mathbf{B}) = 1$

In other words, when learning **B**, we **also** throw out all of the **Ā**-worlds that mapped themselves to a **B̄**-world.

The following question immediately arises: what explains this constraint on rational learning? It seems like it's a semantic constraint (having to do with the meaning of the conditional), not a rational constraint. We might put this point slightly differently as follows. Suppose we spoke a language without conditionals. It seems such a constraint on learning would still apply, and thus we'd be learning a host of conditional facts that we cannot express. This is an odd consequence, but one which we might ultimately find palatable once we see how the view is spelled out.

A second issue raises a possible empirical challenge. Prima facie, you might have thought it intuitively plausible that one could rationally reason via disjunction elimination without having to backtrack and engage in some kind of belief revision. Consider the following line of reasoning:

(42) Either the butler or the gardener did it.

(43) The butler didn't do it.

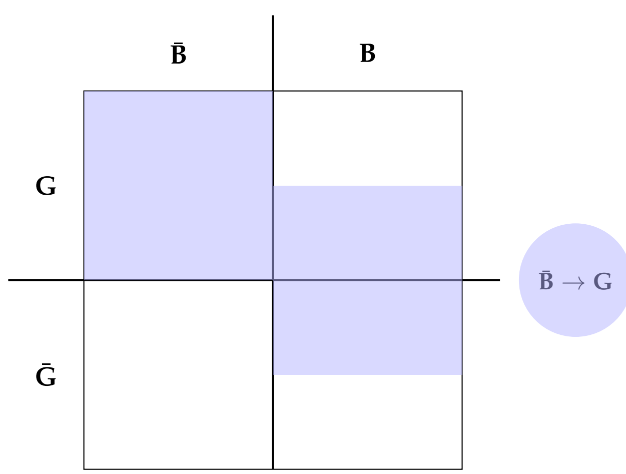
(44) So, the gardener did.

This reasoning seems like straightforward deductive reasoning, requiring only that you eliminate more and more possibilities as you go along. It is unlike, for instance, reasoning involving belief revision, which typically requires some kind of discourse marker indicating that one is revisiting possibilities previously ruled out:

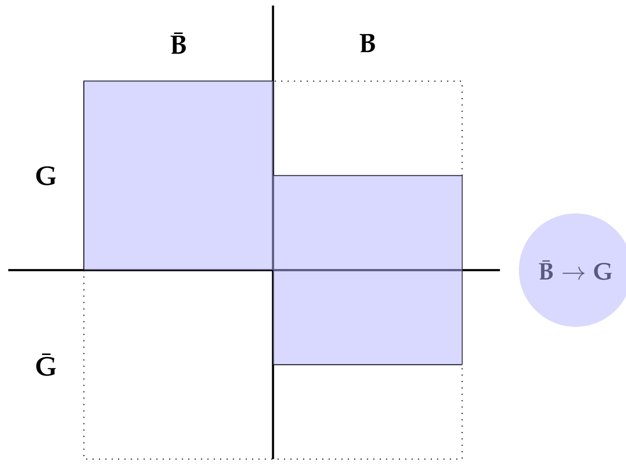
(45) The butler did it.

(46) Wait, actually, maybe not: maybe the butler didn't.

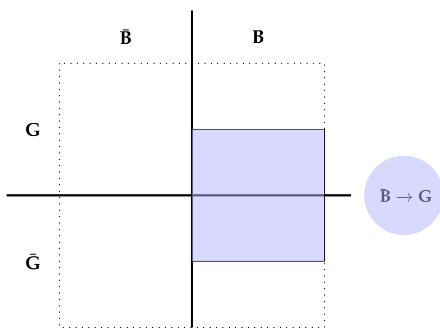
However, on the view under consideration, reasoning via disjunction elimination forces us to engage in belief revision. This is somewhat surprising, so let's go through carefully to see why. Before you learn $B \vee G$, your belief state is:



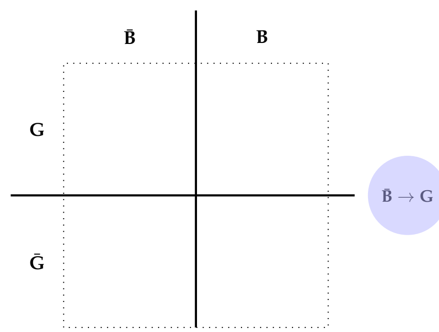
After you learn $\mathbf{B} \vee \mathbf{G}$, you eliminate all of the $\bar{\mathbf{G}}\bar{\mathbf{G}}$-worlds. And in so doing you also eliminate all those **B**-worlds which mapped to a $\bar{\mathbf{G}}$-world, which is sufficient for learning $\bar{\mathbf{B}} \to \mathbf{G}$. Now, your belief state looks like:



Finally, you learn $\bar{\mathbf{B}}$; you must thereby remove any remaining worlds in which $\bar{\mathbf{B}} \to \mathbf{G}$ is true. Otherwise, it could be true even though it has a false presupposition (or think of it this way: there are no available **B**-worlds around to be mapped to). Thus, your resulting belief state is not the only on the left, which reflects merely the learning of $\bar{\mathbf{B}}$, but the one on the right, which is empty!



$P_{i'}(\bar{\mathbf{B}} \to \mathbf{G}) > 0$



$P_{i''}(\bar{\mathbf{B}} \to \mathbf{G}) = 0, P_{i''}(\mathbf{G}) = 0$

Just as with the problems for contextualism, there is room to maneuver here. In particular, once we see the picture fully fleshed out, we will be in a better position to evaluate how unintuitive these consequences really are.

Chapter 5

Tenability

Tenability Results

Van Fraassen's crucial tenability result shows that, given a probability space $\langle W, \mathcal{F}, P\rangle$, given certain assumptions about the space of possibilities over which $A \to B$ is defined, it is possible to ensure that THE THESIS holds for every conditional in $\mathcal{L}_1$ and $\mathcal{L}_2$,

- $\mathcal{L}_0$ is a language consisting of the atomic sentences closed under $\wedge, \vee, \neg$.
- $\mathcal{L}_1 = \mathcal{L}_0 \cup \{\varphi : \exists A, B \in \mathcal{L}_0 : \varphi = A \to B\}$
- $\mathcal{L}_2 = \mathcal{L}_1 \cup \{\varphi : \exists \psi, \chi \in \mathcal{L}_1 : \varphi = \psi \to \chi\}$
  - So, $\mathcal{L}_1$ will contain only conditionals of the form $A \to B$, where A, B are themselves not conditionals.
  - So, $\mathcal{L}_2$ will contain conditionals of the form $A \to B, (A \to B) \to C$, and $A \to (B \to C)$, where A, B are themselves not conditionals (single left- and single right-nesting).

However, there are reasons to wonder whether we should even strive for this level of coverage. We saw before Paolo's counterexample to Stalnaker's Thesis involving right-nested conditionals. And we saw with Stalnaker's proof that there is a trade-off between validating Stalnaker's Thesis for left-nested conditionals and CSO, which we saw is probabilistically valid for simple conditionals. So, maybe all we want is that Stalnaker's Thesis hold for non-nested conditionals. What kinds of commitments must we undertake to ensure this holds?

In light of Hájek (1989)'s wallflower proof, you might think that we require that the range of *Pr* be infinite. But this is in fact not required, so long as we adopt an alternative relationship between *Pr*, which applies to sentences, and *P*, which applies to propositions.

5.1 Semantics and indeterminacy

Here's the basic idea: we start with a Stalnakerian closest-worlds semantics for conditionals. Very often, the "closest" *A*-world will be indeterminate, leading to conditionals to be semantically indeterminate as to what proposition they express. To evaluate the probability of a conditional (or any indeterminate sentence), we take a weighted average of the probabilities of the possible interpretations. A conditional is super true iff it is true on all possible interpretations. Probability is expectation of truth, not super-truth.

Start with the baseline semantics for conditionals (from Stalnaker (1968)):

Semantics
$[\![A \to B]\!]^{i,f,w} = 1$ iff $f(w, \mathbf{A}) \in \mathbf{B}$

Say that an information state i is a pair of a set of worlds s_i and a probability function P_i over subsets of s_i. Let O_i be the set of total permutations of s_i. So, where $s_i = \{w_1, w_2, w_3\}$, we have :

$$O_i = \left\{ \begin{array}{ll} \langle w_1, w_2, w_3 \rangle, & \langle w_1, w_3, w_2 \rangle \\ \langle w_2, w_1, w_3 \rangle, & \langle w_2, w_3, w_2 \rangle \\ \langle w_3, w_1, w_2 \rangle, & \langle w_3, w_2, w_1 \rangle \end{array} \right\}$$

Permutations of worlds correspond to Stalnakerian selection functions satisfying the "indicative constraint" as follows:

Selection functions
f_o is the Stalnakerian selection function corresponding to order $o \in O_i$ iff for every $w \in s_i$:

(a) If $w \in \mathbf{A}$ then $f_o(w, \mathbf{A}) = w$

(b) If $w \notin \mathbf{A}$: then $f_o(w, \mathbf{A})$ = the first $\mathbf{A}$-world in o.

Define the set of admissible selection functions for an information state i, F_i, as the set $\{f_o : o \in O_i\}$. Fixing a particular information state i and $f \in F_i$, the proposition expressed by $A \to B$ is:

$[\![A \to B]\!]^{i,f} = \{w : [\![A \to B]\!]^{i,f,w} = 1\}$, where $f \in F_i$.

Thus, for each non-trivial information state (in which $|F_i| > 1$), there will be multiple candidate selection functions, and hence multiple candidate interpretations of $A \to B$. In such a case, we say that $A \to B$ is semantically indeterminate across a range of propositions:

Indeterminate content
$[\![A]\!]^i = \{\mathbf{X} : \exists f \in F_i : \mathbf{X} = \{w : [\![A]\!]^{i,f,w} = 1\}\}$

This indeterminacy yields a resulting indeterminacy in truth value, which allows us to define a notion of super-truth:

Super-Truth

$$[\![A]\!]^{w,i} = \begin{cases} \text{true} & \text{if } \forall f \in F_i : [\![A]\!]^{i,f,w} = 1 \\ \text{false} & \text{if } \forall f \in F_i : [\![A]\!]^{i,f,w} = 0 \\ \# & \text{otherwise} \end{cases}$$

5.2 Tenability, v1

We now have indeterminacy in our conditionals, so how should we evaluate their probabilities? The original link between Pr and P fails here:

$Pr(A) = P(\mathbf{A})$, where $\mathbf{A}$ is the proposition expressed by A.

One strategy would be to say that sentential probability is probability of supertruth:

$$Pr(A) = P(\{w : \forall f \in F_i\colon [\![A]\!]^{i,f,w} = 1\})$$

But this immediately runs into problems. Consider a simple situation where $s_i = \{w_1, w_2, w_3\}$ and $P_i(\{w_1\}) = P_i(\{w_2\}) = P_i(\{w_3\}) = 1/3$. Let $\mathbf{A} = \{w_1, w_2\}$ and $\mathbf{B} = \{w_2\}$. Then $P_i(\mathbf{B}|\mathbf{A}) = 1/2$. Now consider $\{w : \forall f \in F_i\colon [\![A \to B]\!]^{i,f,w} = 1\}$. This set will contain only w_2, since only w_2 is such that every admissible selection function maps it to an **AB**-world—itself. Consider w_3: some fs will map w_3 and $\mathbf{A}$ to w_1 (a $\bar{\mathbf{B}}$-world) and some will map it and $\mathbf{A}$ to w_2 (a **B**-world). Since not every $f \in F_i$ maps w_3 and $\mathbf{A}$ to a **B**-world, it follows that w_3 is not in our considered set. Thus, it follows that $Pr(A \to B) = 1/3$, thus violating Stalnaker's Thesis.

Instead, what we want is to define sentential probability as expectation of truth, not supertruth. We will assume for now that each interpretation is equally likely (an assumption we'll revisit in the next section). Then, given this, the expectation that A expresses a truth is:

Probability

$$Pr_i(A) = \sum_{f \in F_i} \frac{P_i([\![A]\!]^{i,f})}{|F_i|}$$

Given this account of sentential probability, together with our assumptions about interpretations, we can prove the following:

LIMITED STALNAKER'S THESIS
$Pr_i(A \to B) = Pr_i(B|A)$, if $Pr_i(A) > 0$, for all $A, B \in \mathcal{L}_0$.

5.2.1 The tenability proof

Fact 1. For every $\mathbf{X} \in [\![A \to B]\!]^i$: $\mathbf{AB} \cap s_i \subseteq \mathbf{X}$.

- This holds because of **Selection functions**-(a), the fact that for any $w \in \mathbf{A} \cap s_i$, $f_o(w, \mathbf{A}) = w$, and **Semantics**.

Fact 2. For any $o \in O_i$: if the first **A**-world in o is w, then $f_o(w', \mathbf{A}) = w$, for any $w' \in \bar{\mathbf{A}} \cap s_i$.

- This follows from **Selection functions**-(b): for non-**A**-worlds, the selection function always picks out the same **A**-world—whichever is the first **A**-world in its corresponding sequence.

Fact 3. For each $o \in O_i$ whose first **A**-world is a **B**-world, there is a unique $\mathbf{X} \in [\![A \to B]\!]^i$ such that $\mathbf{X} = s_i \cap (\mathbf{AB} \cup \bar{\mathbf{A}})$. And for $o \in O_i$ whose first **A**-world is a $\bar{\mathbf{B}}$-world, there is a unique $\mathbf{Y} \in [\![A \to B]\!]^i$ such that $\mathbf{Y} = s_i \cap \mathbf{AB}$.

- From **Fact 1**, we know that every $\mathbf{X} \in [\![A \to B]\!]^i$ is true throughout **AB** $\cap s_i$. And from **Fact 2**, we know that every $o \in O_i$ whose **A**-world is a **B**-world will generate a unique **X** which is true throughout $\bar{\mathbf{A}} \cap s_i$ as well. Furthermore, by **Semantics**, we know that **X** will be true nowhere else in s_i. Hence, $\mathbf{X} = s_i \cap (\mathbf{AB} \cup \bar{\mathbf{A}})$.
- And similarly, from **Fact 2**, we know that every $o \in O_i$ whose **A**-world is a **B**-world will generate a unique **Y** which is true nowhere in $\bar{\mathbf{A}} \cap s_i$. Hence, **Y** is true only at $\mathbf{AB} \cap s_i$.
- Since every $o \in O_i$ contains at least one **A**-world, either the first **A**-world in o is a **B**-world or it is a $\bar{\mathbf{B}}$-world. Thus, we can partition the os as to whether their first **A**-world in o is a **B**-world or is a $\bar{\mathbf{B}}$-world, and this partition corresponds to the two-cell partition on propositions in $[\![A \to B]\!]^i$: the $\mathbf{AB} \cup \bar{\mathbf{A}}$-propositions (call this the X-cell), and the **AB**-propositions (call this the X-cell).

Fact 4. The proportion of orders in O_i whose first **A**-world is a **B**-world just is the proportion of **AB**-worlds to **A**-worlds in s_i, and hence equal to $P_i(\mathbf{B}|\mathbf{A})$ (assuming for now that each world in s_i is equally likely to be actual).

- This follows because O_i contains every permutation of worlds in s_i. The scratch calculation goes as follows:
  - Start with the orders whose first world is an **A**-world. Check the proportion of those orders whose first world is a **B**-world. The proportion will be $P_i(\mathbf{B}|\mathbf{A})$.
  - Next, look at the orders whose second world is their first **A**-world. Check the proportion of those orders whose second world is a **B**-world. The proportion will be $P_i(\mathbf{B}|\mathbf{A})$.
  - This pattern will continue until we have reached the point where there are no more **A**-worlds to check. Since at each stage, the proportion of those orders whose first world is a **B**-world is $P_i(\mathbf{B}|\mathbf{A})$, overall the proportion will be $P_i(\mathbf{B}|\mathbf{A})$.

Fact 5. $Pr_i(A \to B) = \sum_{f \in F_i} \frac{Pr_i([\![A \to B]\!]^{i,f})}{|F_i|} = P_i(\mathbf{AB}) \cdot P_i(\bar{\mathbf{B}}|\mathbf{A}) + P_i(\mathbf{AB} \cup \bar{\mathbf{A}}) \cdot P_i(\mathbf{B}|\mathbf{A})$

- It follows from **Fact 3** that $P_i(\mathbf{B}|\mathbf{A})$ of the propositions in $[\![A \to B]\!]^i$ are in the Y-cell (meaning they are equal to $\mathbf{AB} \cup \bar{\mathbf{A}} \cap s_i$), leaving $P_i(\bar{\mathbf{B}}|\mathbf{A})$ of the propositions in the X-cell (meaning they are equal to $\mathbf{AB} \cap s_i$).

We calculate LIMITED STALNAKER'S THESIS as follows:

- Simplifying the right hand side:

1. $P_i(\mathbf{AB}) \cdot P_i(\bar{\mathbf{B}}|\mathbf{A}) + P_i(\mathbf{AB} \cup \bar{\mathbf{A}}) \cdot P_i(\mathbf{B}|\mathbf{A}) =$
2. $P_i(\mathbf{AB}) \cdot P_i(\bar{\mathbf{B}}|\mathbf{A}) + [P_i(\mathbf{AB}) + P_i(\bar{\mathbf{A}})] \cdot P_i(\mathbf{B}|\mathbf{A}) =$
3. $P_i(\mathbf{AB}) \cdot P_i(\bar{\mathbf{B}}|\mathbf{A}) + P_i(\mathbf{AB}) \cdot P_i(\mathbf{B}|\mathbf{A}) + P_i(\bar{\mathbf{A}}) \cdot P_i(\mathbf{B}|\mathbf{A}) =$
4. $P_i(\mathbf{AB}) \cdot \underbrace{[P_i(\bar{\mathbf{B}}|\mathbf{A}) + P_i(\mathbf{B}|\mathbf{A})]}_{1} + P_i(\bar{\mathbf{A}}) \cdot P_i(\mathbf{B}|\mathbf{A}) =$
5. $P_i(\mathbf{AB}) + P_i(\bar{\mathbf{A}}) \cdot P_i(\mathbf{B}|\mathbf{A}) =$
6. $P_i(\mathbf{B}|\mathbf{A}) \cdot P_i(\mathbf{A}) + P_i(\mathbf{B}|\mathbf{A}) \cdot P_i(\bar{\mathbf{A}}) =$
7. $P_i(\mathbf{B}|\mathbf{A})$

5.2.2 Commentary

Let's look at a simple example to see how the system works.

- Let $s_i = \{w_1, w_2, w_3, w_4\}$; $\mathbf{A} = \{w_1, w_2, w_3\}$; $\mathbf{B} = \{w_1, w_2\}$.
- Suppose $P_i(\{w_1\}) = P_i(\{w_2\}) = P_i(\{w_3\}) = P_i(\{w_4\}) = 1/4$.
  - Thus, $Pr_i(B|A) = P_i(\mathbf{B}|\mathbf{A}) = 2/3$
  - Notice also that the probability of any subset of s_i must be a multiple of 1/4, and 2/3 is not a multiple of 1/4. So how can we predict that $Pr_i(A \to B) = 2/3$?
  - Again, the key is that this probability is not the probability of any single proposition, but the expectation of truth throughout a class of propositions, $[\![A \to B]\!]^i$.
- Next, we have:

$$O_i = \left\{ \begin{array}{ll} \langle w_1, w_2, w_3, w_4 \rangle, & \langle w_3, w_1, w_2, w_4 \rangle, \\ \langle w_1, w_2, w_4, w_3 \rangle, & \langle w_3, w_1, w_4, w_2 \rangle, \\ \langle w_1, w_3, w_2, w_4 \rangle, & \langle w_3, w_2, w_1, w_4 \rangle, \\ \langle w_1, w_3, w_4, w_2 \rangle, & \langle w_3, w_2, w_4, w_1 \rangle, \\ \langle w_1, w_4, w_2, w_3 \rangle, & \langle w_3, w_4, w_1, w_2 \rangle, \\ \langle w_1, w_4, w_3, w_2 \rangle, & \langle w_3, w_4, w_2, w_1 \rangle, \\ \langle w_2, w_1, w_3, w_4 \rangle, & \langle w_4, w_1, w_2, w_3 \rangle, \\ \langle w_2, w_1, w_4, w_3 \rangle, & \langle w_4, w_1, w_3, w_2 \rangle, \\ \langle w_2, w_3, w_1, w_4 \rangle, & \langle w_4, w_2, w_1, w_3 \rangle, \\ \langle w_2, w_3, w_4, w_1 \rangle, & \langle w_4, w_2, w_3, w_1 \rangle, \\ \langle w_2, w_4, w_1, w_3 \rangle, & \langle w_4, w_3, w_1, w_2 \rangle, \\ \langle w_2, w_4, w_3, w_1 \rangle, & \langle w_4, w_3, w_2, w_1 \rangle \end{array} \right\}$$

- There are 24 unique orders in O_i, and hence 24 propositions in $[\![A \to B]\!]^i$.
- To calculate the probability of $A \to B$, we must look at the average of the probabilities of each member of $[\![A \to B]\!]^i$.

Step 1:

- Start with the first 12 orders (those whose first world is an **AB**-world, either w_1 or w_2):

$$\begin{array}{c} \langle w_1, w_2, w_3, w_4 \rangle \\ \langle w_1, w_2, w_4, w_3 \rangle \\ \langle w_1, w_3, w_2, w_4 \rangle \\ \langle w_1, w_3, w_4, w_2 \rangle \\ \langle w_1, w_4, w_2, w_3 \rangle \\ \langle w_1, w_4, w_3, w_2 \rangle \\ \langle w_2, w_1, w_3, w_4 \rangle \\ \langle w_2, w_1, w_4, w_3 \rangle \\ \langle w_2, w_3, w_1, w_4 \rangle \\ \langle w_2, w_3, w_4, w_1 \rangle \\ \langle w_2, w_4, w_1, w_3 \rangle \\ \langle w_2, w_4, w_3, w_1 \rangle \end{array}$$

  Each of the corresponding propositions is assigned a probability equal to $P_i(\mathbf{AB} \cup \bar{\mathbf{A}})$.

  - To verify this, take an arbitrary order from this group: $o^* = \langle w_2, w_1, w_4, w_3 \rangle$. The corresponding selection function is f_{o^*}. We know that, for any **A**-world w, $f_{o^*}(w, \mathbf{A}) = w$, and hence:

    At any **AB**-world w': $[\![A \to B]\!]^{i, f_{o^*}, w'} = 1$.

    At any $\mathbf{A}\bar{\mathbf{B}}$-world w'': $[\![A \to B]\!]^{i, f_{o^*}, w''} = 0$.

  - Next, we know that, for any $\bar{\mathbf{A}}$-world w, $f_{o^*}(w, \mathbf{A}) = w_2$. And since w_2 is an **AB**-world, we know that $[\![A \to B]\!]^{i, f_{o^*}, w} = 1$. Hence:

At any $\bar{\mathbf{A}}$-world w''': $[\![A \to B]\!]^{i,f_{o^*},w'''} = 1$.

Step 2:

- Now, consider the next 6 orders (those whose first world is an $\mathbf{A\bar{B}}$-world, w_3:

$$\langle w_3, w_1, w_2, w_4 \rangle$$
$$\langle w_3, w_1, w_4, w_2 \rangle$$
$$\langle w_3, w_2, w_1, w_4 \rangle$$
$$\langle w_3, w_2, w_4, w_1 \rangle$$
$$\langle w_3, w_4, w_1, w_2 \rangle$$
$$\langle w_3, w_4, w_2, w_1 \rangle$$

  Each of the corresponding propositions is assigned a probability equal to $P_i(\mathbf{AB})$.

  - To verify this, take an arbitrary order from this group: $o^{**} = \langle w_3, w_4, w_1, w_2 \rangle$. The corresponding selection function is $f_{o^{**}}$. We know that, for any $\mathbf{A}$-world w, $f_{o^{**}}(w, \mathbf{A}) = w$, and hence:

    At any $\mathbf{AB}$-world w': $[\![A \to B]\!]^{i,f_{o^{**}},w'} = 1$.

    At any $\mathbf{A\bar{B}}$-world w'': $[\![A \to B]\!]^{i,f_{o^{**}},w''} = 0$.

  - Next, we know that, for any $\bar{\mathbf{A}}$-world w, $f_{o^{**}}(w, \mathbf{A}) = w_3$. And since w_3 is an $\mathbf{A\bar{B}}$-world, we know that $[\![A \to B]\!]^{i,f_{o^{**}},w} = 0$. Hence:

    At any $\bar{\mathbf{A}}$-world w''': $[\![A \to B]\!]^{i,f_{o^{**}},w'''} = 0$.

Step 3:

- Now consider the next 4 orders (those whose first world is a $\bar{\mathbf{A}}$-world and whose second is an $\mathbf{AB}$-world):

$$\langle w_4, w_1, w_2, w_3 \rangle$$
$$\langle w_4, w_1, w_3, w_2 \rangle$$
$$\langle w_4, w_2, w_1, w_3 \rangle$$
$$\langle w_4, w_2, w_3, w_1 \rangle$$

  Each of the corresponding propositions is assigned a probability equal to $P_i(\mathbf{AB} \cup \bar{\mathbf{A}})$. The reasoning is the same as at step 1 above.

Step 4:

- Now consider the last 2 orders (those whose first world is a $\bar{\mathbf{A}}$-world and whose second is an $\mathbf{A\bar{B}}$-world):

$$\langle w_4, w_3, w_1, w_2 \rangle$$
$$\langle w_4, w_3, w_2, w_1 \rangle$$

Each of the corresponding propositions is assigned a probability equal to $P_i(\mathbf{AB})$. The reasoning is the same as at step 2 above.

Step 5:

- We now calculate the probability of $A \to B$ by summing as follows:

$$Pr_i(A \to B) = \sum_{f \in F_i} \frac{Pr_i([\![A \to B]\!]^{i,f})}{|F_i|}$$

$$= P_i(\mathbf{AB} \cup \bar{\mathbf{A}}) \cdot 1/2 + P_i(\mathbf{AB}) \cdot 1/4 + P_i(\mathbf{AB} \cup \bar{\mathbf{A}}) \cdot 1/6 + P_i(\mathbf{AB}) \cdot 1/12$$
$$= 3/4 \cdot 1/2 + 1/2 \cdot 1/4 + 3/4 \cdot 1/6 + 1/2 \cdot 1/12$$
$$= 3/8 + 1/8 + 3/24 + 1/24$$
$$= 1/2 + 4/24$$
$$= 2/3$$

Notice that this result holds even in a finite model, demonstrating that the limitation of Hájek's wallflower proof. That proof only works if you have the simple relationship between *Pr* and *P*, which does not hold here because of how we incorporate the indeterminacy of conditionals into their probabilities.

There are two crucial limitations to this result. One is that A, B must themselves not be conditionals. This is because $A \to B$ does not determine a set of worlds, but rather a set of propositions, so, compositionally, we have no natural way of calculating either:

$$(A \to B) \to C$$
$$A \to (B \to C)$$

As discussed before, we think this is actually a good limitation of the result: we do not actually want Stalnaker's Thesis to apply to these cases. The second limitation is the assumption that each world in s_i is equally likely to be actual. What happens if we relax this assumption?

5.3 Generalizing

If we relax the assumption that each world in s_i is equally likely to be actual, then our prior calculation will fail to yield LIMITED STALNAKER'S THESIS. Roughly, this is because counting up the proportion of interpretations at which $A \to B$ is true will no longer track the probabilities of its component parts.

To overcome this limitation, we need to generalize **Probability** so that it takes a weighted average on the interpretations, rather than a simple average. But that means we need to calculate a weight on the orders. How are we to do this?

- Start with a probability mass function p_i over worlds, which we can appeal to to define P_i as follows:

$$P_i(\mathbf{A}) = \sum_{w \in \mathbf{A}} p_i(w)$$

- Let $o[w, \ldots, w_n]$ be the set of os exactly like $\langle o, \ldots, o_n \rangle$ through its nth member.

- We then define a probability measure over subsets of orders, Q_i, as follows:

(q1) $Q_i(O_i) = 1$

(q2) $Q_i(o[w]) = \dfrac{p_i(w)}{P_i(s_i)}$

(q3) $Q_i(o[w, w']) = \dfrac{p_i(w')}{P_i(s_i \setminus \{w\})} \cdot Q_i(o[w])$

(q4) $Q_i(o[w, \ldots, w_n]) = \dfrac{p_i(w_n)}{P_i(s_i \setminus \{w, \ldots, w'\})} \cdot Q_i(o[w, \ldots, w_{n-1}])$

- Roughly, what Q_i does is weight each order by the corresponding weights on its worlds. If you think about the orders as chances of picking worlds without replacement, this will ensure that at each stage of selection, you are more likely to pick higher-weighted worlds than non-higher-weighted worlds.

Next, we define Pr_i in terms of p_i and Q_i:

Probability*
$Pr_i(A) = \sum\limits_{w} \sum\limits_{o} p_i(w) \cdot Q_i(\{o\}) \cdot [\![A]\!]^{w,i,f_o}$

- This supplants **Probability**, and it takes into consideration the probabilities over orders we have just defined.

From here, we can establish:

Fact 4*. For any $o \in O_i$: the probability that the first **A**-world of o is a **B**-world just is $P_i(\mathbf{B}|\mathbf{A})$.

- Start with some notation:

- Let $O_i(n, \mathbf{A})$ be the set of orders in O_i whose first **A**-world is their nth.

- Let $O_i(n, \mathbf{B}|\mathbf{A})$ be the set of orders in O_i whose first **A**-world is their nth and also a **B**-world.

We will prove **Fact*** by proving a few Lemmas:

Lemma 1. $Q_i(O_i(1,\mathbf{B}|\mathbf{A})|O_i(1,\mathbf{A})) = P_i(\mathbf{B}|\mathbf{A})$

Proof. This states that the probability of picking an order whose first **A**-world is its 1st and also a **B**-world is equal to $P_i(\mathbf{B}|\mathbf{A})$.

We know that:

(i) $Q_i(O_i(1,\mathbf{A})) = \sum_{w\in\mathbf{A}} Q_i(o[w]) = \sum_{w\in\ \mathbf{A}} p_i(w)$

(ii) $Q_i(O_i(1,\ \mathbf{B}|\ \mathbf{A})) = \sum_{w\in\ \mathbf{AB}} Q_i(o[w]) = \sum_{w\in\ \mathbf{AB}} p_i(w)$

Thus, since Q_i is a probability function, we have:

$$Q_i(O_i(1,\ \mathbf{B}|\ \mathbf{A})|O_i(1,\ \mathbf{A})) = \frac{Q_i(O_i(1,\mathbf{B}|\mathbf{A}))}{Q_i(O_i(1,\mathbf{A}))}$$

And thus,

$$\frac{Q_i(O_i(1,\mathbf{B}|\mathbf{A}))}{Q_i(O_i(1,\mathbf{A}))} = \frac{\sum_{w\in\mathbf{AB}} p_i(w)}{\sum_{w\in\mathbf{A}} p_i(w)} = \frac{P_i(\mathbf{AB})}{P_i(\mathbf{A})}$$

Let's extend this reasoning one level: calculating the probability that the second world is an **AB**-world, given that the first is a $\bar{\mathbf{A}}$-world and the second is an **A**-world.

Lemma 2. $Q_i(O_i(2,\mathbf{B}|\mathbf{A})|O_i(2,\mathbf{A})) = P_i(\mathbf{B}|\mathbf{A})$

Proof. Start with the following facts:

(i) $Q_i(O_i(2,\mathbf{A})) = \sum_{w\in\bar{\mathbf{A}}} \sum_{w'\in\mathbf{A}} Q_i(o[w,w']) = \sum_{w\in\bar{\mathbf{A}}} \sum_{w'\in\mathbf{A}} \frac{p_i(w')}{P_i(s_i \setminus \{w\})} \cdot p_i(w)$

(ii) $Q_i(O_i(2,\mathbf{B}|\mathbf{A})) = \sum_{w\in\bar{\mathbf{A}}} \sum_{w'\in\mathbf{AB}} Q_i(o[w,w']) = \sum_{w\in\bar{\mathbf{A}}} \sum_{w'\in\mathbf{AB}} \frac{p_i(w')}{P_i(s_i \setminus \{w\})} \cdot p_i(w)$

As above, we have:

$$Q_i(O_i(2,\mathbf{B}|\mathbf{A})|O_i(2,\mathbf{A})) = \frac{Q_i(O_i(2,\mathbf{B}|\mathbf{A}))}{Q_i(O_i(2,\mathbf{A}))}$$

Hence,

$$\frac{Q_i(O_i(2,\mathbf{B}|\mathbf{A}))}{Q_i(O_i(2,\mathbf{A}))} = \frac{\sum_{w\in\bar{\mathbf{A}}}\sum_{w'\in\mathbf{AB}} \frac{p_i(w')}{P_i(s_i\setminus\{w\})}\cdot p_i(w)}{\sum_{w\in\bar{\mathbf{A}}}\sum_{w'\in\mathbf{A}} \frac{p_i(w')}{P_i(s_i\setminus\{w\})}\cdot p_i(w)} =$$

$$\frac{\sum_{w\in\bar{\mathbf{A}}}\sum_{w'\in\mathbf{AB}} \frac{p_i(w')}{P_i(s_i\setminus\{w\})}}{\sum_{w\in\bar{\mathbf{A}}}\sum_{w'\in\mathbf{A}} \frac{p_i(w')}{P_i(s_i\setminus\{w\})}} = \sum_{w\in\bar{\mathbf{A}}}\sum_{w'\in\mathbf{AB}} \frac{p_i(w')}{P_i(s_i\setminus\{w\})}\cdot \sum_{w\in\bar{\mathbf{A}}}\sum_{w'\in\mathbf{A}} \frac{P_i(s_i\setminus\{w\})}{p_i(w')} =$$

$$\sum_{w\in\mathbf{AB}}\sum_{w\in\mathbf{A}} \frac{p_i(w)}{p_i(w)} = \frac{P_i(\mathbf{AB})}{P_i(\mathbf{A})}$$

The crucial move is to extend this reasoning up to $n = 1 + |\bar{\mathbf{A}}|$:

Lemma 3. For any $n \leq 1 + |\bar{\mathbf{A}}|$: $Q_i(O_i(n,\mathbf{B}|\mathbf{A})|O_i(n,\mathbf{A})) = P_i(\mathbf{B}|\mathbf{A})$.

- Let $n \leq 1 + |\bar{\mathbf{A}}|$. Start with the following facts again:

  (i) $Q_i(O_i(n,\mathbf{A})) = \sum_{w,\ldots,w'\in\bar{\mathbf{A}}}\sum_{w_n\in\mathbf{A}} Q_i(o[w,\ldots,w',w_n]) =$

  $$\sum_{w,\ldots,w'\in\bar{\mathbf{A}}}\sum_{w_n\in\mathbf{A}} \frac{p_i(w_n)}{P_i(s_i\setminus\{w,\ldots,w'\})}\cdot Q_i(o[w,\ldots,w'])$$

  (ii) $Q_i(O_i(n,\mathbf{B}|\mathbf{A})) = \sum_{w,\ldots,w'\in\bar{\mathbf{A}}}\sum_{w_n\in\mathbf{AB}} Q_i(o[w,\ldots,w',w_n]) =$

  $$\sum_{w,\ldots,w'\in\bar{\mathbf{A}}}\sum_{w_n\in\mathbf{AB}} \frac{p_i(w_n)}{P_i(s_i\setminus\{w,\ldots,w'\})}\cdot Q_i(o[w,\ldots,w'])$$

- As above, we have:

  $$Q_i(O_i(n,\mathbf{B}|\mathbf{A})|O_i(n,\mathbf{A})) = \frac{Q_i(O_i(n,\mathbf{B}|\mathbf{A}))}{Q_i(O_i(n,\mathbf{A}))}$$

- Hence,

$$\frac{Q_i(O_i(n,\mathbf{B}|\mathbf{A}))}{Q_i(O_i(n,\mathbf{A}))} = \frac{\sum_{w,\ldots,w'\in\bar{\mathbf{A}}}\sum_{w_n\in\mathbf{AB}} \frac{p_i(w_n)}{P_i(s_i\setminus\{w,\ldots,w'\})}\cdot Q_i(o[w,\ldots,w'])}{\sum_{w,\ldots,w'\in\bar{\mathbf{A}}}\sum_{w_n\in\mathbf{A}} \frac{p_i(w_n)}{P_i(s_i\setminus\{w,\ldots,w'\})}\cdot Q_i(o[w,\ldots,w'])} =$$

$$\frac{\sum\limits_{w,\ldots,w'\in\bar{\mathbf{A}}}\ \sum\limits_{w_n\in\mathbf{AB}} \dfrac{p_i(w_n)}{P_i(s_i\setminus\{w,\ldots,w'\})}}{\sum\limits_{w,\ldots,w'\in\bar{\mathbf{A}}}\ \sum\limits_{w_n\in\mathbf{A}} \dfrac{p_i(w_n)}{P_i(s_i\setminus\{w,\ldots,w'\})}} =$$

$$\sum_{w,\ldots,w'\in\bar{\mathbf{A}}}\ \sum_{w_n\in\mathbf{AB}} \frac{p_i(w_n)}{P_i(s_i\setminus\{w,\ldots,w'\})} \cdot \sum_{w,\ldots,w'\in\bar{\mathbf{A}}}\ \sum_{w_n\in\mathbf{A}} \frac{P_i(s_i\setminus\{w,\ldots,w'\})}{p_i(w_n)} =$$

$$\sum_{w\in\mathbf{AB}}\ \sum_{w\in\mathbf{A}} \frac{p_i(w)}{p_i(w)} = \frac{P_i(\mathbf{AB})}{P_i(\mathbf{A})}$$

Thus, we know, for each $n \leq 1 + |\bar{\mathbf{A}}|$, that the probability that the nth world of an order is an **AB**-world, given that the $n-1$th is a $\bar{\mathbf{A}}$-world and the nth is an **A**-world is equal to $P_i(\mathbf{B}|\mathbf{A})$. Let $O_i^{\mathbf{AB}}$ be the set of orders in O_i whose first **A**-world is a **B**-world. It follows from **Lemma 3** that $Q_i(O_i^{\mathbf{AB}}) = P_i(\mathbf{B}|\mathbf{A})$, which establishes **Fact 4***.

Finally, from **Fact 2** and **Semantics**, it follows that the only positive instances of the following equation are $o \in O_i^{\mathbf{AB}}$:

$$\sum_{w\in\bar{\mathbf{A}}}\ \sum_{o} p_i(w)\cdot Q_i(\{o\})\cdot [\![A \to B]\!]^{i,f_o,w}$$

This, together with **Fact 4***, yields:

Fact 5. $\sum\limits_{w\in\bar{\mathbf{A}}}\ \sum\limits_{o} p_i(w)\cdot Q_i(\{o\})\cdot [\![A \to B]\!]^{i,f_o,w} = \sum\limits_{w\in\bar{\mathbf{A}}} p_i(w)\cdot P_i(\mathbf{B}|\mathbf{A}) = P_i(\bar{\mathbf{A}})\cdot P_i(\mathbf{B}|\mathbf{A})$

From Total Probability we have:

(*) $Pr_i(A\to B) = Pr_i(A\to B|A)\cdot Pr_i(A) + Pr_i(A\to B|\bar{A})\cdot Pr_i(\bar{A})$

Then, given **Semantics**-(a), we have:

$$Pr_i(A\to B|A) = P_i(\mathbf{B}|\mathbf{A})$$

And given **Probability***, we have:

$$Pr_i(A\to B|\bar{A}) = \frac{\sum\limits_{w\in\bar{\mathbf{A}}}\ \sum\limits_{o} p_i(w)\cdot Q_i(\{o\})\cdot [\![A \to B]\!]^{w,i,f_o}}{P_i(\bar{\mathbf{A}})}$$

But then from **Fact 5**, it follows from this that:

$$Pr_i(A\to B|\bar{A}) = \frac{P_i(\mathbf{B}|\mathbf{A})\cdot P_i(\bar{\mathbf{A}})}{P_i(\bar{\mathbf{A}})} = P_i(\mathbf{B}|\mathbf{A})$$

So, plugging back into (*) yields LIMITED STALNAKER'S THESIS:

$$Pr_i(A \to B) = \underbrace{Pr_i(A \to B|A)}_{P_i(\mathbf{B}|\mathbf{A})} \cdot Pr_i(A) + \underbrace{Pr_i(A \to B|\bar{A})}_{P_i(\ \mathbf{B}|\ \mathbf{A})} \cdot Pr_i(\bar{A}) = P_i(\ \mathbf{B}|\ \mathbf{A})$$

Chapter 6

Empirical Failures of the Thesis

We have supposed throughout that Stalnaker's Thesis is valid for every non-conditional antecedent/consequent, at a single probability function. However, we now come to revisit this assumption. Counterexamples to even this limited version of Stalnaker's Thesis have been proposed by McGee (2000) and Kaufmann (2004). We discuss such counterexamples in this chapter, and suggest a way of amending the Stalnakerian semantics above to predict them.

6.1 The counterexamples

Red Ticket. Smith just drew a ticket from one of two boxes (assigned randomly). In Box 1, there were 100 tickets, 90 red, 81 red with a dot. In Box 2, there were 100 tickets, 10 red, 1 red with a dot. How likely is it that:

(47) If Smith drew a red ticket, it had a dot.

There seem to be two, equally natural, lines of reasoning about the probability of (47):

- *One-half*. He either drew from Box 1 or Box 2. If he drew from Box 1, then he had a 9 in 10 chance of drawing a red ticket with a dot, assuming he drew a red ticket. If he drew from Box 2, then he had a 1 in 10 chance of drawing a red ticket with a dot, assuming he drew a red ticket. Since he was equally likely to draw from either box, the probability that if he drew a red ticket, it had a dot, should be 1/2 (the average of these probabilities).
  - This answer violates Stalnaker's Thesis, since the conditional probability that the ticket Smith drew had a dot, given that he drew a red ticket is high (see reasoning below).
- *High*. Suppose Smith drew a red ticket. Then it's much more likely that he drew from Box 1 than Box 2, since the proportion of red tickets was greater in Box 1 than Box 2. But then, given that he drew from Box 1, he had a 9 in 10 chance of drawing a red ticket with a dot, assuming he drew a red ticket. So, it is highly likely that if Smith drew a red ticket, it had a dot.

Cases of this kind were first proposed by McGee (2000); Kaufmann (2004). If these intuitions are correct, then (47) constitutes a simple counterexample to Stalnaker's Thesis (on at least one of its interpretations, more on that below).

Let's consider another case, just to be sure this is not some fluke:

> **Coin Flip**. A coin has just been tossed. You know it was weighted 3 to 1 towards tails, so it probably landed on tails. Jones knows how the coin landed, but you couldn't hear what she said; you think it's likely that she said, "it landed heads." Thus, you conclude that it is likely that:
>
> (48) If Jones is correct, the coin landed heads.
>
> Nonetheless, it seems you should think the following is unlikely:
>
> (49) If Jones is correct, she said that the coin landed heads.
>
> This is because you still think it likely that the coin landed on tails (since it was weighted towards tails). So, it is more likely that Jones said that it landed tails if she was correct.

Together, the intuitions that (48) is likely but (49) is unlikely constitute a counterexample to Stalnaker's Thesis. The reason is that, conditional on Jones being correct, "the coin landed heads" is true iff "Jones said that the coin landed heads" is true. Thus, at least one of (48) or (49) must be a counterexample to Stalnaker's Thesis.

6.2 Diagnosing

Recall that Stalnaker's Thesis follows from two principles:

- PROBABILITY CENTERING: $Pr(A \to B|A) = Pr(B|A)$
- INDEPENDENCE: $Pr(A \to B|A) = Pr(A \to B)$

So, violations of Stalnaker's Thesis must involve a failure of one or both of these principles. Which is most plausible? Well, Probability Centering seems to be valid for all indicative conditionals. It follows from Strong Centering, which is confirmed by betting intuitions:

STRONG CENTERING: $\models A \supset [B \equiv (A \to B)]$

If we bet on $A \to B$, then I win if AB holds, and you win if $A\bar{B}$ holds (set aside what happens for now if A fails to hold). This is reason to think that AB is sufficient for the truth of $A \to B$, and $A\bar{B}$ sufficient for the falsity of $A \to B$, thus evidence favoring Strong Centering, and hence Probability Centering.

Furthermore, it does seem plausible that the counterexamples above are violations of Independence. The one-half probability of $R \to D$ in **Red Ticket** does not hold were one to learn R. For, if one learned R, then one would think it much more likely that Smith drew from Box 1 than Box 2, and thus come to think it likely that his ticket had a dot. Thus, it seems that the counterexamples involve failures of Independence.

It follows, then, that we get violations of Stalnaker's Thesis just if $Pr(A \to B|\bar{A}) \neq Pr(B|A)$. This means that all of the action will happen at worlds where the conditional's antecedent is false.

6.3 Local and Global

Kaufmann (2004) suggests that conditionals have two interpretations—a global interpretation that validates Restricted Stalnaker's Thesis, and a local interpretation that allows for violations of Restricted Stalnaker's Thesis:

Global

$$Pr(A \to B) = \sum_{\mathbf{Z} \in \mathbb{Z}} P(\mathbf{B}|\mathbf{AZ}) \cdot P(\mathbf{Z}|\mathbf{A}) = P(\mathbf{B}|\mathbf{A})$$

Local

$$Pr(A \to_{\mathbb{Z}} B) = \sum_{\mathbf{Z} \in \mathbb{Z}} P(\mathbf{B}|\mathbf{AZ}) \cdot P(\mathbf{Z})$$

These come apart iff for some $\mathbf{Z} \in \mathbb{Z}$, $P(\mathbf{Z}) \neq P(\mathbf{Z}|\mathbf{A})$.

Kaufmann's proposal represents a major advance in how to think of these violations of Stalnaker's Thesis. However, it raises several questions of its own:

- Is this an ambiguity?

  Kaufmann is unclear (see Kaufmann (2004): 603 for some evidence that he thinks it is an ambiguity, however, see Kaufmann (2009): 26-27 for evidence that he is thinking of the difference purely pragmatically). Recall that, for someone like Edgington who bakes something like Stalnaker's Thesis into the meaning of *if*, either denying the intuitions or opting for ambiguity seem to be the only options.

- What kind of semantics predicts these two interpretations?

  Kaufmann does not say, although Kaufmann (2005) sketches a semantic theory that could be integrated with the discussion here to provide a compositional theory. We will sketch the beginnings of such a theory below.

- What is $\mathbb{Z}$?

  Kaufmann tentatively suggests that it is an objective chance determining variable that is causally independent of, but evidential dependent on, A. Although see Khoo (2016) for discussion of whether this is always so. Khoo suggests that the partition is determined by a contextually salient question under discussion.

6.4 Generalizing Stalnaker

We can generalize Stalnaker's semantics in a natural way to predict these violations of Stalnaker's Thesis without positing ambiguity. To see why ambiguity is unnecessary, notice that if conditionals always have the local interpretation, then they will still have a Thesis-friendly interpretation whenever they are interpreted relative to a partition $\mathbb{Z}$ such that for all $\mathbf{Z} \in \mathbb{Z}$, $P(\mathbf{Z}) = P(\mathbf{Z}|\mathbf{A})$. One natural such partition is the trivial one: $\mathbb{Z} = \{s_i\}$.

Start with the basic Stalnaker semantics from earlier:

> **Semantics**
> $[\![A \to B]\!]^{i,f,w} = 1$ iff $f(w, \mathbf{A}) \in \mathbf{B}$

Recall that we defined a set of orders corresponding to all of the permutations of worlds in the relevant information state s_i. This corresponds to Stalnaker's "indicative constraint":

> "I cannot define the selection function in terms of the context set, but the following constraint imposed by the context on the selection function seems plausible: if the conditional is being evaluated at a world in the context set, then the world selected must, if possible, be within the context set as well (where C is the context set, if $i \in C$, then $f(A, i) \in C$). In other words, all worlds within the context set are closer to each other than any worlds outside it. The idea is that when a speaker says 'If A,' then everything he is presupposing to hold in the actual situation is presupposed to hold in the hypothetical situation in which A is true." (Stalnaker (1975): 275-6)

The idea is simple: if you start within s_i, then you stay within s_i when looking for the closest **A**-world. The set of permutations of worlds in s_i corresponded to all the possible ways of finding an **A**-world in s_i, from any starting point (this reflected the fact that the context is indeterminate in which respects of closeness mattered).

Following Khoo (2016), we are going to introduce a way of constraining the indeterminacy of closeness relations in the context by a partition. The new constraint implements the idea behind Stalnaker's indicative constraint at the level of a cell of this partition:

> **Partition Constraint**
> For any cell of $\mathbf{Z} \in \mathbb{Z}$, if $w \in \mathbf{Z}$ and $\mathbf{A} \cap \mathbf{Z} \neq \emptyset$, then $f_{\mathbb{Z}}(w, \mathbf{A}) \in \mathbf{Z}$.

What this says is that, if you start in cell **Z**, then the selection function must select an **A**-world from within **Z**. Thus, this constraint just is a generalization of Stalnaker's indicative constraint at the level of the cells of the relevant partition.

Now, it should be clear from the tenability result above that, as long as each cell $\mathbf{Z} \in \mathbb{Z}$ is such that $\mathbf{Z} \cap \mathbf{A} \neq \emptyset$, then $P_i(\mathbf{A} \to \mathbf{B}|\mathbf{Z}) = P_i(\mathbf{B}|\mathbf{AZ})$. So, the probability of the conditional just is its probability at each cell of $\mathbb{Z}$, as weighted by the probability of that cell. Thus, we immediately predict Kaufmann's local probability for conditionals:

$$Pr_i(A \rightarrow_{\mathbb{Z}} B) = \sum_{\mathbf{Z} \in \mathbb{Z}} P(\mathbf{B}|\mathbf{AZ}) \cdot P(\mathbf{Z})$$

In order to establish the result, we still need to show how to lift the **Partition Constraint** into a constraint on the orders in O_i, and we also need to generalize our definition of Pr_i to account for the fact that different worlds are constrained by different cells. In what follows, we briefly sketch one way of doing this (we leave fleshing out the proposal in full detail for future work).

First, we restate the **Partition Constraint** as a global constraint on the possible interpretations of $A \rightarrow_{\mathbb{Z}} B$ as follows. First, we define $O_i^{\mathbb{Z}}$ as a function from cells of $\mathbb{Z}$ to sets of permutations of the worlds in that cell. So, where $s_i = \{w_1, w_2, w_3\}$ and $\mathbb{Z} = \{\{w_1, w_2\}, \{w_3\}\}$, we have:

$$O_i^{\mathbb{Z}}(\{w_1, w_2\}) = \{\langle w_1, w_2\rangle, \langle w_2, w_1\rangle\}$$
$$O_i^{\mathbb{Z}}(\{w_3\}) = \{\langle w_3\rangle\}$$

We can now define, for each cell $\mathbf{Z} \in \mathbb{Z}$, the set of admissible Stalnakerian selection functions for that cell, $F_i^{\mathbb{Z}}(\mathbf{Z})$ in terms of $O_i^{\mathbb{Z}}(\mathbf{Z})$ as follows:

Partitional Selection functions
f_o is the Stalnakerian selection function corresponding to order $o \in O_i^{\mathbb{Z}}(\mathbf{Z})$, for some $\mathbf{Z} \in \mathbb{Z}$ iff for every $w \in s_i \cap \mathbf{Z}$:

(a) If $w \in \mathbf{A}$ then $f_o(w, \mathbf{A}) = w$

(b) If $w \notin \mathbf{A}$: then $f_o(w, \mathbf{A})$ = the first $\mathbf{A}$-world in o.

- $F_i^{\mathbb{Z}}(\mathbf{Z})$ is the set of selection functions corresponding to orders in $O_i^{\mathbb{Z}}(\mathbf{Z})$

We now need to introduce a definedness condition on $A \rightarrow B$ for f, w: this is to ensure that we don't combine selection functions corresponding to orders that don't contain w with w.

$[\![A \rightarrow_{\mathbb{Z}} B]\!]^{f_o,w}$ is defined only if $o \in O_i^{\mathbb{Z}}([w]_{\mathbb{Z}})$; where $[w]_{\mathbb{Z}}$ is the cell of $\mathbb{Z}$ that w falls within.

- Let $\mathbf{Def}_i^w(A)$ = the set of orders $o \in O_i$ for which $[\![A]\!]^{f_o,w}$ is defined.
- Then, $\mathbf{Def}_i^w(A \rightarrow_{\mathbb{Z}} B) = O_i^{\mathbb{Z}}([w]_{\mathbb{Z}})$

Let's look at an example to see how this works. Suppose $\mathbf{A} = \{w_2, w_3\}, \mathbf{B} = \{w_2\}$, and, as above, $\mathbb{Z} = \{\{w_1, w_2\}, \{w_3\}\}$. We then have the following:

- $F_i^{\mathbb{Z}}(\{w_1, w_2\}) = f_1$, where for any $w \in s_i$: $f_1(w, \mathbf{A}) = w_2$.

- $F_i^{\mathbb{Z}}(\{w_3\}) = f_2$, such that $f_2(w_3, \mathbf{A}) = w_3$.
- Hence,

  $[\![A \rightarrow_{\mathbb{Z}} B]\!]^{f_1,w_1} = 1$, since $f_1(w_1, \mathbf{A}) \in \mathbf{B}$

  $[\![A \rightarrow_{\mathbb{Z}} B]\!]^{f_1,w_2} = 1$, since $f_1(w_2, \mathbf{A}) \in \mathbf{B}$

  $[\![A \rightarrow_{\mathbb{Z}} B]\!]^{f_2,w_3} = 1$, since $f_2(w_3, \mathbf{A}) \in \mathbf{B}$

  All other combinations of f, w are undefined for $A \rightarrow_{\mathbb{Z}} B$.
- Hence, we predict that $A \rightarrow_{\mathbb{Z}} B$ has probability 1, even though $P(\mathbf{B}|\mathbf{A}) = 1/2$.

To achieve this last result, we need to modify **Probability*** slightly to take into account the new definedness constraint:

Probability* (Partitional)
$$Pr_i(A) = \sum_{w} \sum_{o \in \mathbf{Def}_i^w(A)} p_i(w) \cdot Q_i(\{o\}|\mathbf{Def}_i^w(A)) \cdot [\![A]\!]^{w,i,f_o}$$

Given this, we then predict that $Pr(A \rightarrow_{\mathbb{Z}} B) = 1$, even though $Pr(B|A) = 1/2$:

$$Pr_i(A \rightarrow B) =$$
$$\sum_{w} \sum_{o \in \mathbf{Def}_i^w(A)} p_i(w) \cdot Q_i(\{o\}|\mathbf{Def}_i^w(A)) \cdot [\![A]\!]^{w,i,f_o} =$$
$$\sum_{w} \sum_{o \in O_i^{\mathbb{Z}}([w]_{\mathbb{Z}})} p_i(w) \cdot Q_i(\{o\}|O_i^{\mathbb{Z}}([w]_{\mathbb{Z}})) \cdot [\![A]\!]^{w,i,f_o} =$$
$$\underbrace{p_i(w_1)}_{1/3} \cdot \underbrace{Q_i(\{f_1\}|O_i^{\mathbb{Z}}([w_1]_{\mathbb{Z}}))}_{1} \cdot 1 + \underbrace{p_i(w_2)}_{1/3} \cdot \underbrace{Q_i(\{f_1\}|O_i^{\mathbb{Z}}([w_2]_{\mathbb{Z}}))}_{1} \cdot 1 + \underbrace{p_i(w_3)}_{1/3} \cdot \underbrace{Q_i(\{f_2\}|O_i^{\mathbb{Z}}([w_3]_{\mathbb{Z}}))}_{1} \cdot 1 =$$
$$1/3 + 1/3 + 1/3 = 1$$

This is just to illustrate how the theory works. We leave aside the proofs for the fully general system for future work.

Chapter 7

Triviality for Counterfactuals

The interaction between countefactuals and probability is a topic worthy of a whole course in itself. But here we can provide a quick overview of a couple of triviality results for counterfactuals. The first is presented in Robert and Williams 2012, while the second is an unpublished result in Santorio 2018. Both result from modifications of standard triviality results for indicatives.

7.1 Williams' result

Williams' (2012) central idea is that nothing in Lewis's standard triviality proof relies specifically on the relevant conditionals being indicatives. So we can reproduce it for counterfactuals, with some tweaks. The overall strategy can be summed up in four steps:

i. We start from a counterfactual counterpart of the Thesis, which connects credences in counterfactuals with a notion of suppositional credence.

ii. From here, we derive a Thesis-like principle about chance.

iii. We assume (and argue for) a closure assumption about chance.

iv. Run a Lewis-like triviality result using the assumptions in (i)–(iii).

7.1.1 Suppositional credence

Williams assumes a notion of *suppositional credence*. This is the credence that subjects assigns to a proposition, given a certain supposition. As an example of a suppositional credence, consider the credence that you assign to the proposition that Kennedy would have been assassinated, on the supposition that Oswald had not shot him (presumably, a low credence). Notice that suppositional credences are a set of attitudes you have in addition to your standard credences. Suppositional credences come both in the indicative and counterfactual variety; following Williams, we use '$Pr^A(C)$' to denote a subject's suppositional credence in C, on the supposition that A.

Once we have a notion of suppositional credence in the background, we can state some coordination principles between ordinary credences in conditionals and suppositional credences. For the purposes of this argument, we are concerned with the counterfactual version of this principle. Using the standard '$\Box\!\!\rightarrow$' symbol to denote counterfactuals:

Counterfactual Ramsey Test (CRT). $Pr(A \Box\!\!\rightarrow C) = Pr^A(C)$

The full strength of (CRT) is not needed to run the argument. As Williams points out, we get a triviality results also using principles that are weaker:

Counterfactual Ramsey Bound. $Pr(A \Box\!\!\rightarrow C) \leq Pr^A(C)$

Counterfactual Ramsey Zero. If $Pr^A(C) = 0$, then $Pr(A \Box\!\!\rightarrow C) = 0$

For convenience, throughout our discussion we'll be sticking to (CRT).

As stated, (CRT) is not a very substantial constraint, since we haven't said anything about the notion of counterfactual suppositional credence. (CRT) becomes more substantial once we adopt a second constraint on counterfactual suppositional credence, which links the latter to chance (for the origins of this constraint, see Skyrms 1980b).

Skyrms' Thesis. $Pr^A(C) = \sum_{w_i \in W} Pr(w_i) \times Ch_{w_i}(C \mid A)$

Skryms' Thesis is sometimes taken to be an analog of Stalnaker's Thesis for counterfactuals. Informally, it says that one's counterfactual suppositional credence in C on the supposition that A should be identical to one's expectation of the conditional chance of C, given A.

As Williams notices, Skyrms' Thesis is a conditional version of Lewis's so-called Principal Principle, i.e. a principle, originally formulated by David Lewis (1981), connecting credence and chance.[1] The appeal to chance requires two qualifications, which we relegate to a footnote.[2]

7.1.2 Deriving counterfactual triviality

At this point, we have all the assumptions we need to derive a triviality result. Assume first that we have an agent who is ideally informed about the chance function. For this agent, Skyrms' Thesis reduces to the simple equation:

Informed Skyrms' Thesis. $Pr^A(C) = Ch(C \mid A)$

[1] Here a simplified formulation of the Principal Principle:

Principal Principle. $Pr(A \mid Ch(A) = x) = x$, provided that $Pr(Ch(A) = x) > 0$

Informally, the Principal Principle says that, conditional on the information that the chance of A is x, one's credence in A should also be x. Even more informally: the Principal Principle says that an agent's credences should align with the known chances.

[2] First, we need to deal with so-called inadmissible information. It seems that chances should constrain credences only for agents that have no inadmissible information, i.e. only for agents that have no access to information about the future, such as information they might gather from crystal balls or the like. Second, chances evolve over time. So we will need to index the chance function to appropriate temporal indices.

Moreover, we assume the original Principal Principle:

Principal Principle. $Pr(A \mathrel{\Box\!\!\rightarrow} C) = Ch(A \mathrel{\Box\!\!\rightarrow} C)$

Putting these together with the Counterfactual Ramsey test ($Pr^A(C) = Pr(A \mathrel{\Box\!\!\rightarrow} C)$), we get

Chancy Equation. $Ch(C \mid A) = Ch(A \mathrel{\Box\!\!\rightarrow} C)$

At this point we're almost ready to run a triviality proof. We only need the usual assumption that the relevant probabilities (this time, chances) are closed under conditionalization:

> **Chance closure**: For any X and Ch: if $Ch(\cdot)$ models a possible chance distribution, then $Ch(\cdot|X)$ models a possible chance distribution.

At this point, we run an argument that is exactly analogous to Lewis's original triviality result, only using chances instead of credences. We assume that Ch'

i. $Ch(A \mathrel{\Box\!\!\rightarrow} C) =$
ii. $Ch(A \mathrel{\Box\!\!\rightarrow} C \mid C) \times Ch(C) + Ch(A \mathrel{\Box\!\!\rightarrow} C \mid \neg C) \times Ch(\neg C) =$
iii. $Ch'(A \mathrel{\Box\!\!\rightarrow} C) \times Ch(C) + Ch''(A \mathrel{\Box\!\!\rightarrow} C) \times Ch(\neg C) =$
iv. $Ch'(C \mid A) \times Ch(C) + Ch''(C \mid A) \times Ch(\neg C) =$
v. $1 \times Ch(C) + 0 \times Ch(\neg C) =$
vi. $Ch(C)$

As usual, this result is not acceptable. Resistance strategies will involve challenging one or more of the assumptions that go into the proof. The options involve challenging (CRT), challenging the Closure Principle, and challenging the Principal Principle (for the latter strategy, see Schwarz 2016). Discussion of these options goes beyond the purposes of these notes.

7.2 A collapse result for counterfactuals

The next result we examine (from Santorio 2018) is a collapse result for *would*-counterfactuals. Given seemingly plausible assumptions, we can prove that $A \mathrel{\Box\!\!\rightarrow} C$ and $A \mathrel{\Diamond\!\!\rightarrow} C$ have the same probability—i.e., that a *would*-counterfactual and the corresponding *might*-counterfactual have the same probability.

Collapse. $Pr(A \mathrel{\Box\!\!\rightarrow} C) = Pr(A \mathrel{\Diamond\!\!\rightarrow} C)$

This result is obviously implausible.

We offer two proofs of Collapse. Both exploit on a notion of counterfactual suppositional credence, similar to the one used by Williams. In addition, both appeal to the following principles about counterfactual suppositional credence:

Suppositional Excluded Middle (SEM). $Pr^A(C) + Pr^A(\neg C) = 1$

Counterfactual Ramsey Test (CRT). $Pr(A \mathrel{\Box\!\!\rightarrow} C) = Pr^A(C)$

(SEM) should be uncontroversial. It is part and parcel of any existing construal of suppositional credence, and it is an immediate consequence of the claim that subjunctive credences are probabilistic. So it seems pretty safe. (CRT) is familiar from the discussion of Williams' result. Notice that, differently from Williams, we're not assuming any further constraint of the notion of counterfactual suppositional credence. We are just assuming that, whatever suppositional credence turn out to be, a subject's counterfactual suppositional credence equals the credence in the relevant counterfactual.

The two proofs differ with respect to the other assumptions they use. Let us examine them in detail.

7.2.1 Duality and the first collapse result

For the first result, we assume:

Duality. $(A \,\Box\!\!\rightarrow C) \dashv\vDash \neg(A \,\Diamond\!\!\rightarrow \neg C)$

Let us flag right away that Duality is an implausible assumption in this context. There is a well-known tension between Duality and Conditional Excluded Middle[3] But it's useful to go through this proof, since the second proof will derive a probabilistic version of Duality, starting from much more plausible premises.

The proof is very simple:

i. $Pr^A(C) + Pr^A(\neg C) = 1$ (SEM)
ii. $Pr(A \,\Box\!\!\rightarrow C) + Pr(A \,\Box\!\!\rightarrow \neg C) = 1$ (i, CRT)
iii. $Pr(A \,\Box\!\!\rightarrow C) + Pr(\neg(A \,\Diamond\!\!\rightarrow C)) = 1$ (ii, Duality)
iv. $Pr(A \,\Box\!\!\rightarrow C) + 1 - Pr(A \,\Diamond\!\!\rightarrow C) = 1$ (iii, PC)
v. $Pr(A \,\Box\!\!\rightarrow C) = Pr(A \,\Diamond\!\!\rightarrow C)$ (iv, algebra)

This result is not too surprising. It is often remarked that counterfactual semantics that vindicate Duality interact poorly with probability (see e.g. Edgington 2008, Schulz 2014a). What is surprising is that the same conclusion can be derived *without* Duality, and that hence the problem potentially generalizes to all kinds of counterfactual semantics—including those that are supposed to interact better with probability, such as e.g. Stalnaker's semantics.

7.2.2 Assumptions for the second collapse result

The strategy of the second result is simple: we use two principles about probabilities of conditionals to derive a probabilistic counterpart of Duality.

Probabilistic Duality. (PD) $Pr(A \,\Box\!\!\rightarrow C) = Pr(\neg(A \,\Diamond\!\!\rightarrow \neg C))$

Then we use (PD) to run exactly the same proof we gave in the previous section.

We use two assumptions about probabilities of counterfactuals:

[3]Specifically: if we assume classical logic, Duality and Conditional Excluded Middle lead to the logical equivalence of *would*- and *might*-counterfactuals.

Nonzero. For all Pr such that $Pr(A \diamond\!\!\rightarrow C) > 0$, $Pr(A \Box\!\!\rightarrow C \mid A \diamond\!\!\rightarrow C) > 0$

Upper bound. If $Pr(\neg(A \diamond\!\!\rightarrow \neg C)) = 1$, then $Pr(A \Box\!\!\rightarrow C) = 1$

We also appeal to a principle of conditional logic:

Conditional Non-Contradiction (CNC) $A \Box\!\!\rightarrow \neg C \vDash \neg(A \Box\!\!\rightarrow C)$

Finally, we appeal to the usual closure assumption for credence, which is familiar from other triviality proofs in these notes.

(CNC) is a basic principle of conditional logic, so it should not be in question. Let me say something to motivate Nonzero and Upper Bound.

Nonzero says that the probability of *If A, would C*, conditional on *If A, might C*, has to be greater than zero. It captures the intuition that it seems irrational to be certain of (1), and yet assign zero credence to (2).

(50) If Sarah had tossed the coin, it might have landed tails.

(51) If Sarah had tossed the coin, it would have landed tails.

Nonzero can be questioned by appealing to the idea that propositions that express live possibilities can still receive probability zero (see e.g. Hájek 2003). In particular, we might grant that, in some cases, a *might*-counterfactual is true while the corresponding *would*-counterfactual has probability zero. As a candidate example, consider:

(52) If Sarah picked a real number at random between 0 and 1, she might pick 0.5.

(53) If Sarah picked a real number at random between 0 and 1, she would pick 0.5.

This is a real concern. But, rather than trying to argue against it, we can sidestep it. We grant that **Nonzero** might have limited applicability. We will still be able to derive the collapse result for a large subclass of counterfactuals: 50 and 51 are exactly cases of this sort. This is bad enough.

Upper Bound says that, if you are certain of the negation of *If A, might C*, then you should be certain of *If A, would* $\neg C$. Given a classical treatment of negation, this is the probabilistic counterpart of one direction of Duality, namely:

Not-might*-to-*if $\neg(A \diamond\!\!\rightarrow C) \vDash (A \Box\!\!\rightarrow \neg C)$

We are *not* assuming **Duality** for the purposes of this proof. But we think that theorists of all stripes should be happy with *Not-might*-to-*if*, and hence with Upper Bound. Notice that *Not-might*-to-*if* can be seen as saying that Lewis-style truth-conditions for counterfactuals entail the truth conditions of natural language counterfactuals. Now, theorists who depart from Lewis invariably complain that his semantics is too strong; but it is hard to deny that Lewis's truth conditions entail the truth conditions of counterfactuals. The controversial direction of Duality is the left-to-right one, which we are not assuming here.

7.2.3 Proof of the second collapse result

Step 1: incompatibility of $A \;\Box\!\!\rightarrow \neg C$ and $A \;\Diamond\!\!\rightarrow C$. The first step establishes that a counterfactual $A \;\Box\!\!\rightarrow \neg C$ and the *might*-counterfactual $A \;\Diamond\!\!\rightarrow C$ are incompatible: i.e., their conjunction has probability zero. Assume for *reductio* that $A \;\Box\!\!\rightarrow \neg C$ and $A \;\Diamond\!\!\rightarrow C$ are compatible and that hence some probability function Pr assigns positive probability to both of them. Via Nonzero, we know:

i. $Pr(A \;\Box\!\!\rightarrow C \mid A \;\Diamond\!\!\rightarrow C) > 0$

Assuming that the class of rational credence functions is closed under conditionalization, we have:

ii. $Pr_{A \Box\!\!\rightarrow \neg C}(A \;\Box\!\!\rightarrow C \mid A \;\Diamond\!\!\rightarrow C) > 0$

Via the definition of conditionalization, (ii) is equivalent to:

iii. $Pr(A \;\Box\!\!\rightarrow C \mid A \;\Diamond\!\!\rightarrow C \wedge A \;\Box\!\!\rightarrow \neg C) > 0$

However, via **CNC**, we know that

iv. $Pr(A \;\Box\!\!\rightarrow C \mid A \;\Box\!\!\rightarrow \neg C) = 0$

Hence (iii) and (iv) contradict. We conclude that $A \;\Box\!\!\rightarrow \neg C$ and $A \;\Diamond\!\!\rightarrow C$ are incompatible.

Step 2: equivalence of $A \;\Box\!\!\rightarrow \neg C$ and $\neg(A \;\Diamond\!\!\rightarrow \neg C)$. Take any Pr such that $Pr(\neg(A \;\Diamond\!\!\rightarrow C)) > 0$. Then we can derive that $Pr(A \;\Box\!\!\rightarrow C)$ is equal to $Pr(\neg(A \;\Diamond\!\!\rightarrow \neg C))$. We first observe, via total probability:

i. $Pr(A \;\Box\!\!\rightarrow \neg C) = Pr(A \;\Box\!\!\rightarrow \neg C \wedge A \;\Diamond\!\!\rightarrow C) + Pr(A \;\Box\!\!\rightarrow \neg C \wedge \neg(A \;\Diamond\!\!\rightarrow C))$

Via the previous proof, $Pr(A \;\Box\!\!\rightarrow \neg C \wedge A \;\Diamond\!\!\rightarrow C) = 0$. Reorganizing the term on the right-hand side:

ii. $Pr(A \;\Box\!\!\rightarrow \neg C) = Pr(A \;\Box\!\!\rightarrow \neg C \mid \neg(A \;\Diamond\!\!\rightarrow C)) \times Pr(\neg(A \;\Diamond\!\!\rightarrow C))$

Via the closure condition, $Pr(\cdot \mid \neg(A \;\Diamond\!\!\rightarrow C))$ is a rational credence function. Since $Pr(\neg(A \;\Diamond\!\!\rightarrow C) \mid \neg(A \;\Diamond\!\!\rightarrow C)) = 1$, via **Upper Bound** we get that $Pr(A \;\Box\!\!\rightarrow \neg C) \mid \neg(A \;\Diamond\!\!\rightarrow C)) = 1$. Hence (ii) simplifies to

iii. $Pr(A \;\Box\!\!\rightarrow \neg C) = Pr(\neg(A \;\Diamond\!\!\rightarrow C))$

Assuming that negation is classical, we get:

> **Probabilistic Duality.** $Pr(A \;\Box\!\!\rightarrow C) = Pr(\neg(A \;\Diamond\!\!\rightarrow \neg C))$

At this point, we can run the same proof as in §7.2.1 to prove that $A \;\Box\!\!\rightarrow C$ and $A \;\Diamond\!\!\rightarrow C$ have the same probability.

As usual, this result is not acceptable and resistance strategies will involve challenging one or more of the assumptions that go into the proof. The options involve challenging (CRT), challenging the Closure Principle,

Bibliography

Ernest Adams. The logic of conditionals. *Inquiry*, 8(1):166–197, 1965.

Ernest Adams. *The Logic of Conditionals*, volume 86 of *Synthese Library*. D. Reidel, Boston, 1975.

Ernest W. Adams. *The Logic of Conditionals: An Application of Probability to Deductive Logic*. D. Reidel Pub. Co., 1996.

Andrew Bacon. Stalnaker's thesis in context. *The Review of Symbolic Logic*, 8(1):131–163, 2015.

Jonathan Bennett. *A Philosophical Guide to Conditionals*. Oxford University Press, Oxford, 2003.

Justin Bledin. Modus ponens defended. *Journal of Philosophy*, 112(2):57–83, 2015.

Richard Bradley. A preservation condition for conditionals. *Analysis*, 60(267):219–222, 2000.

Richard Bradley. A defence of the ramsey test. *Mind*, 116(461):1–21, 2007.

Nate Charlow. Triviality for restrictor conditionals. *Noûs*, 50(3):533–564, 2016.

Dorothy Edgington. On conditionals. *Mind*, 104:235–329, 1995.

Dorothy Edgington. Counterfactuals. *Proceedings of the Aristotelian Society*, 108:1–21, 2008.

Branden Fitelson. The strongest possible lewisian triviality result. *Thought: A Journal of Philosophy*, 4(2):69–74, 2015.

Branden Fitelson. Two new(ish) triviality results for the indicative conditional. Lecture Notes, 2016. URL `http://fitelson.org/triviality`$_h$`andout.pdf`.

Anthony Gillies. On truth-conditions for if (but not quite only if). *Philosophical Review*, 118 (3):325–349, 2009.

Anthony Gillies. Iffiness. *Semantics & Pragmatics*, 3:1–42, 2010. doi: 10.3765/sp.3.4.

Alan Hájek. Probabilities of conditionals - revisited. *Journal of Philosophical Logic*, 18:423–428, 1989.

Alan Hájek. What conditional probability could not be. *Synthese*, 137(3):273–323, 2003.

Alan Hájek. Triviality pursuit. *Topoi*, 30:3–15, 2011a.

Alan Hájek. Triviality pursuit. *Topoi*, 30:3–15, 2011b.

Frank Jackson. On assertion and indicative conditionals. *Philosophical Review*, 88:565–89, 1979.

Frank Jackson. *Conditionals*. Blackwell, Oxford, 1987.

Stefan Kaufmann. Conditioning against the grain: Abduction and indicative conditionals. *Journal of Philosophical Logic*, 33:583–606, 2004.

Stefan Kaufmann. Conditional predictions. *Linguistics and Philosophy*, 28:181–231, 2005.

Stefan Kaufmann. Conditionals right and left: Probabilities for the whole family. *Journal of Philosophical Logic*, 38:1–53, 2009.

Justin Khoo. A note on gibbard's proof. *Philosophical Studies*, 166(1):153–164, 2013.

Justin Khoo. Probability of conditionals in context. *Linguistics and Philosophy*, 39(1):1–43, 2016. doi: 10.1007/s10988-015-9182-z.

Justin Khoo. Quasi indexicals. *Philosophy and Phenomenological Research*, 2018.

Justin Khoo and Matthew Mandelkern. Triviality results and the relationship between logical and natural languages. *Mind*, forthcoming.

Peter Klecha. Two kinds of sobel sequences: Precision in conditionals. In Ulrike Steindl, Thomas Borer, Huilin Fang, Alfredo Garca Pardo, Peter Guekguezian, Brian Hsu, Charlie OâĂŹHara, and Iris Chuoying Ouyang, editors, *Proceedings of WCCFL 32*, pages 131–140. Cascadilla Press, 2015.

Angelika Kratzer. The notional category of modality. In H. J. Eikmeyer and H. Rieser, editors, *Words, Worlds, and Contexts: New Approaches to Word Semantics*. de Gruyter, Berlin, 1981.

Angelika Kratzer. Conditionals. *Chicago Linguistics Society*, 22(2):1–15, 1986.

Angelika Kratzer. Modality. *Semantics: An international handbook of contemporary research*, pages 639–650, 1991.

Angelika Kratzer. *Modals and Conditionals: New and Revised Perspectives*, volume 36. Oxford University Press, 2012.

David Lewis. Adverbs of quantification. In Edward L. Keenan, editor, *Formal Semantics of Natural Language*. Cambridge University Press, 1975.

David Lewis. Probabilities of conditionals and conditional probabilities. *The Philosophical Review*, 85:297–315, 1976. ISSN 00318108. doi: 10.2307/2184045. URL http://links.jstor.org/sici?sici=0031-8108%28197607%2985%3A3%3C297%3APOCACP%3E2.0.CO%3B2-T.

David Lewis. *IFS: Conditionals, Belief, Decision, Chance and Time*, chapter A Subjectivist's Guide to Objective Chance, pages 267–297. Springer Netherlands, Dordrecht, 1981. ISBN 978-94-009-9117-0. doi: 10.1007/978-94-009-9117-$0_1$4. URL http://dx.doi.org/10.1007/978-94-009-9117-$0_1$4.

David Lewis. Probabilities of conditionals and conditional probabilities ii. *The Philosophical Review*, 95:581–589, October 1986. ISSN 00318108. URL http://links.jstor.org/sici?sici=0031-8108%28198610%2995%3A4%3C581%3APOCACP%3E2.0.CO%3B2-D.

Vann McGee. A counterexample to modus ponens. *Journal of Philosophy*, 82(9):462–471, 1985.

Vann McGee. To tell the truth about conditionals. *Analysis*, 60(1):107–111, 2000.

Peter Milne. The simplest lewis-style triviality proof yet? *Analysis*, 63(4):300–303, 2003.

J. Robert and G. Williams. Counterfactual triviality. *Philosophy and Phenomenological Research*, 85(3):648–670, 2012.

Daniel Rothschild. Do indicative conditionals express propositions? *Noûs*, 47(1):49–68, 2013.

Jeffrey Sanford Russell and John Hawthorne. General dynamic triviality theorems. *Philosophical Review*, 125(3):307–339, 2016.

Paolo Santorio. Collapse for counterfactual probabilities. draft, University of California, San Diego, 2018.

Paolo Santorio and J. Robert.G. Williams. Triviality for indeterminate languages. draft, University of California, San Diego and University of Leeds, 2018.

Moritz Schulz. Counterfactuals and arbitrariness. *Mind*, 123(492):1021–1055, 2014a.

Moritz Schulz. Counterfactuals and arbitrariness. *Mind*, 123(492):1021–1055, 2014b.

Wolfgang Schwarz. Subjunctive conditional probability. forthcoming in *Journal of Philosophical Logic*, 2016.

Brian Skyrms. The prior propensity account of subjunctive conditionals. In William Harper, Robert C. Stalnaker, and Glenn Pearce, editors, *Ifs*. Reidel, 1980b.

Robert Stalnaker. A theory of conditionals. In N. Recher, editor, *Studies in Logical Theory*. Oxford, 1968.

Robert Stalnaker. Indicative conditionals. *Philosophia*, 5, 1975.

Robert Stalnaker. Letter to van fraassen. In W. Harper and C. Hooker, editors, *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, volume 1, pages 302–306. Reidel, Dordrecht, 1976.

Frank Veltman. Defaults in update semantics. *Journal of Philosophical Logic*, 25(3):221–261, 1996.

J. Robert G. Williams. Conversation and conditionals. *Philosophical Studies*, 138(2):211–223, 2008.

Seth Yalcin. Epistemic modals. *Mind*, 116(464):983–1026, 2007.

Seth Yalcin. Probability operators. *Philosophy Compass*, 5(11):916–937, 2010.