Conditional predictions

A probabilistic account

Stefan Kaufmann (kaufmann@northwestern.edu)
Northwestern University

Abstract. The connection between the probabilities of conditionals and the corresponding conditional probabilities has long been explored in the philosophical literature, but its implementation faces both technical obstacles and objections on empirical grounds. In this paper I first outline the motivation for the probabilistic turn and Lewis' triviality results, which stand in the way of what would seem to be its most straightforward implementation. I then focus on Richard Jeffrey's 'random-variable' approach, which circumvents these problems by giving up the notion that conditionals denote propositions in the usual sense. Even so, however, the random-variable approach makes counterintuitive predictions in simple cases of embedded conditionals. I propose to address this problem by enriching the model with an explicit representation of causal dependencies. The addition of such causal information not only remedies the shortcomings of Jeffrey's conditional, but also opens up the possibility of a unified probabilistic account of indicative and counterfactual conditionals.

Keywords: conditionals, counterfactuals, probability, causality

Introduction

Most current theories of conditionals are inspired by Ramsey's (1929) paraphrase of the process involved in their evaluation:

(RT) If two people are arguing 'If $p$ will $q$?' and are both in doubt as to $p$, they are adding $p$ hypothetically to their stock of knowledge and arguing on that basis about $q$ ... We can say they are fixing their degrees of belief in $q$ given $p$.

This suggestion can be made precise in a number of ways, depending on what is assumed about beliefs and belief update. One often-made pair of assumptions is the following:

i. Degrees of belief are measured by probabilities.

ii. The hypothetical addition of the antecedent proceeds by condition-alization.

With these assumptions, the Ramsey Test comes down to the following doctrine, which I will refer to as "the Thesis":\(^1\)

\(^1\) Also known as "Stahlacker's Thesis" after Stahlacker (1970) (Stahlacker, 1976 disavowed it); "Adams' Thesis" after Adams (1965, 1975); or the "Conditional
(T) The probability of a conditional ‘if A then C’ is the conditional probability of C, given A.

All of these notions will be explained in more detail below. As we will see, (T) as stated here is still too general and requires further elaborations regarding the interpretation of the probability measure and the exact role of the conditional probability. I am going to propose a detailed account for the case of simple and right-nested predictive conditionals based on (T) and discuss its potential as a unified account of predictive conditionals and their counterfactual counterparts. Further extensions, such as an application to “epistemic” conditionals (see below) and conditionals with conditional antecedents, are possible, but cannot be accommodated within this paper and are left for future occasions.

The probabilistic approach, its motivation, its merits and its specific problems in connection to conditionals have all been studied extensively, but this work went largely unnoticed in the linguistic literature (but see Cohen, 2003 for some discussion). Therefore a considerable part of this paper is devoted to introducing the various strands that will come together later on.

In Section 1 I explain my choice to deal with one class of indicative conditionals at the exclusion of others. Some of the merits of the probabilistic account and its relationship to the more familiar quantificational framework are discussed in Section 2. In Section 3 I introduce the basic definitions of the probabilistic framework and the relationship between truth and probability for truth-functional sentences. I then briefly discuss why the same strategy is not available for conditionals if (T) is to hold in general, and how this problem can be avoided by making the denotation of the conditional dependent upon the probability distribution. This move is motivated and formalized in Section 4 on the relationship between truth and chance in time. In Sections 5 and 6 I discuss the predictions that this account makes about the probabilities of right-nested indicative conditionals and counterfactuals, and argue that reference to causal dependencies is required to correct certain counterintuitive claims. The effect of such causal information in the interpretation of conditionals is discussed in Section 7. Finally, in Section 8 I end with an outlook of some further issues not addressed in this paper.

Construal of Conditional Probability (CCCP)” following Hájek and Hall (1994) and Hall (1994).
1. Classification

I will focus on predictive conditionals and their counterfactual counterparts, illustrated in (1a,b), respectively. I will not deal with the class of epistemic conditionals (1c).

(1) a. If you strike the match, it will light. [predictive]
   b. If you had struck the match, it would have lit. [counterfactual]
   c. If you struck the match, it lit. [epistemic]

Within the wider landscape of conditionals, all three of (1a–c) are direct and specific in the terms of Quirk et al. (1985).

The decision to limit the scope of this paper to predictive and counterfactual conditionals at the exclusion of epistemic ones is not motivated by any assumption of a fundamental semantic difference between these classes. On the contrary, I believe that the treatment I will propose for predictive conditionals is just as applicable, mutatis mutandis, to epistemic ones. What differs is the interpretation of the probability measure, roughly corresponding to the difference between various accessibility relations, or “modal bases,” in non-probabilistic approaches. In this paper I will focus on objective or “metaphysical” readings and the attendant interpretation of probability as objective chance. It is this limitation which places epistemic conditionals beyond the scope of the paper, since an adequate account of the latter must involve reference to epistemic uncertainty. A unified treatment, which must include an account of the interaction between the two interpretations, has to wait for another occasion.

Even granting that there are good reasons for limiting coverage in the interest of readability, however, my restricting it to what under some analyses is not considered a natural class of conditionals deserves some explanation. In the remainder of this section, I will give some relevant arguments.

1.1. Indicative vs. Counterfactual

According to one classification, (1a,c) are indicative and (1b) is counterfactual. Subjunctive is sometimes used instead of the latter and usually

---

2 Such a fundamental difference has been postulated in connection with otherwise somewhat similar proposals; see for instance Gibbard’s (1981) distinction between “nearness” and “epistemic” readings.
meant to be coextensive with it. A similar line is drawn between “open” and “hypothetical” conditionals by Quirk et al. (1985).

This classification has wide currency despite the well-known fact that both “counterfactual” and “subjunctive” are misnomers. Anderson’s (1951) example (2) shows that counterfactuals can be used to argue for the truth of the antecedent (see also Stalnaker, 1975; Adams, 1976; Karttunen and Peters, 1979; Barwise, 1986; Comrie, 1986; von Fintel, 1998).

(2) If he had taken arsenic, he would have shown just these symptoms [those which he in fact shows]

“Subjunctive,” as a morphological category, is useless in drawing the distinction. The use of the subjunctive mood in conditionals is optional, restricted to formal style, and possible in sentences of both types (Quirk et al., 1985; Dudman, 1988; Bennett, 1988; Edgington, 1995). Thus (3a) is a subjunctive non-counterfactual whereas (3c) (unlike 3b) is an indicative counterfactual.4

(3) a. If any person be found guilty, he shall have the right to appeal.
   b. If I were rich, I would buy you anything you wanted.
   c. If I was rich, I would buy you anything you wanted.

More pertinent than these terminological issues, however, is the question of how deep a semantic division ought to be drawn between the two classes. The minimal pair in (4), due to Adams (1970), is often cited in this connection.

(4) a. If Oswald did not kill Kennedy, someone else did.
   b. If Oswald had not killed Kennedy, someone else would have.

From the observation that “[4a] is probably true while [4b] may very well be false,” Lewis (1973) concludes that “there really are two different sorts of conditional; not a single conditional that can appear as indicative or as counterfactual depending on the speaker’s opinion about the truth of the antecedent.” (p. 3)

Not everyone agrees. Strawson (1986) presents a pair like the following (using different sentences):

---
3 Opinions vary on sentences like ‘If you were to strike the match, it would light’, which Lewis (1973) claims are subjunctives that “appear to have the truth conditions of indicative conditionals.” (p. 4)
4 One reviewer questioned the grammaticality of (3c), but many informants do find it acceptable.
(5)  a. Remark made on November 21, 1963:
    “If Oswald does not kill Kennedy, someone else will.”

   b. Remark made on November 23, 1963:
    “If Oswald had not killed Kennedy, someone else would have.”

   In Strawson’s opinion, “[i]t seems obvious that about the least attractive thing one could say about the difference between these two remarks is that it shows that, or even that it is partly accounted for by the fact that, the expression ‘if ... then ...’ has a different meaning in one remark from the meaning which it has in the other.” (p. 230) Edgington (1995) agrees and suggests that the difference may be “more like the difference between mature cheddar and freshly-made cheddar than the difference between chalk and cheese.” (p. 239)

   Notice that (4a) is an epistemic conditional whereas (5a) is predictive. It is perhaps no coincidence that the pairs in (4a,b) and (5a,b) have been used to argue against and for a unified account, respectively. Both of (5a,b) are conditional predictions made from the perspective of different times, referring not merely to (non-actual) facts but to alternative courses of events, which depart from the actual history at a point in time prior to that referred to by the antecedent.5 This opinion is held by many authors and implemented in a variety of accounts (Downing, 1959; Adams, 1975; Ellis, 1978; Thomason and Gupta, 1981; Tedeschi, 1981; Dudman, 1994; Strawson, 1986; Bennett, 1988; Mellor, 1993; Edgington, 1995; Dahl, 1997; Dancygier, 1998, and others).

   Aside from the intuitions that motivate these unified accounts, the close semantic affinity between counterfactuals and indicatives is demonstrated, perhaps even more forcefully, by the fact that they exhibit very similar behavior in inference. In particular, the well-known inadequacies of the material and strict conditionals—the fact that they validate patterns to which there is ample linguistic counter-evidence—can be demonstrated equally well with examples from either class (Veltman, 1985). In Section 2.1 I will illustrate some of these facts with indicative examples, but the reader can easily verify that a substitution with their respective counterfactual counterparts yields similar judgments. But if inference patterns are to be the semanticist’s primary source of empirical data, then the claim that this parallelism arises in spite of a deep semantic difference is hard to substantiate.

---

5 The fact that the departure from the actual course of events lies in the past from the perspective of the reference time of the antecedent has been exploited by Dudman, Dahl and others to give an explanation of backshift, the addition of a “layer” of past morphology in the antecedents of counterfactuals. For alternative views, see James (1982), Fleischman (1989), Iatrion (2000).
1.2. Predictive vs. Non-predictive

An alternative classification draws the major dividing line between predictive and counterfactual conditionals on the one hand and epistemic ones, on the other (Dancygier, 1998). Dudman (1984, 1994 and elsewhere) motivates this claim with the observation that while the antecedent of (1c) retains its interpretation when used in isolation as in (6b), that of (1a), when used as in (6a), is only felicitous under a "scheduling" interpretation (cf. also Palmer, 1983; Comrie, 1985).

(6)  a. You strike the match.
    b. You struck the match.

Funk (1985) arrives at a similar distinction on different grounds, relating the difference between (1a) and (1c) to one in the status of the facts that determine the truth or falsehood of the antecedent: They are not yet "manifested" in the case of (1a), and "manifested" but unknown in the case of (1c). In Funk's words, "the meaning of the conditioning frame can be said to vary from 'if it happens that...’ to 'if it is true that ...’ " (p. 376) He goes on to discuss examples from languages in which the distinction has clearer morphological reflexes than it does in English.

The relevant difference, as I see it, is directly related to the reasons I gave above for excluding epistemic conditionals from this paper. Indicative conditionals are generally used under uncertainty as to the truth of the antecedent: The addressee may or may not strike (1a) or have struck (1c) the match. The distinction Funk appeals to concerns the nature of that uncertainty. In the epistemic conditional (1c), it is purely subjective: The speaker is ignorant of the relevant facts, but objectively, those facts are settled. In the case of the predictive conditional (1a), on the other hand, the world itself may yet follow different future courses of events, and whether the addressee eventually does or doesn't strike the match cannot be determined until the facts are in. The uncertainty is objective.

In general, the interaction between these kinds of uncertainty is a one-way affair: Objective uncertainty implies subjective uncertainty under the reasonable assumption that what is objectively not yet determined cannot already be known. On the other hand, many settled facts, for instance that the match was not struck at the time in question, remain unknown to most epistemic subjects.

This difference has clearly felt consequences for the question of whether there is something objectively "correct" to be believed about the conditional. It seems appropriate to say that (7a) is true or false

kaufmann@northwestern.edu
Draft - do not cite. Comments welcome.
July 6, 2003
depending on whether the match is dry or wet, and likewise in retrospect for the corresponding counterfactual (7b). In these cases, the correctness of a speaker’s belief about the conditionals can be judged against this objective standard—the beliefs of an “expert,” as it were, who knows all the settled relevant facts.

(7) a. If the match is struck, it will light.
   b. If the match had been struck, it would have lit.
   c. If the match was struck, it lit.

Things are not so clear for the epistemic conditional (7c). If the match was not struck, there does not seem to be anything objectively “correct” to be believed about it. Nor do other facts, such as whether the match was dry or not, have any bearing on the objective value of (7c). The “expert” falls silent. However, a speaker who knows that the match was wet but does not know whether it was struck may still consistently believe that the conditional is false. More generally, Gibbard (1981) shows that when the antecedent is false, it is possible for two speakers with different sets of correct but partial beliefs to arrive at diametrically opposed conclusions about the conditional, neither of which is proven wrong by the facts (although from the fact that they disagree one can infer that the antecedent is false). Edgington (1995) shows that this possibility arises not only with past reference, but whenever the antecedent at speech time has no (objective) chance of turning out true.

These facts suggest that in a unified account of conditionals, the two kinds of uncertainty should be kept separate and some care must be taken in setting up the relationship between them. In focusing on predictive conditionals, I can avoid the complexities that this would require. To summarize, two assumptions will be made throughout the rest of the paper: (i) The probabilities are uniformly interpreted as objective chance, and (ii) the antecedents of the conditional discussed have some chance of turning out true.

2. Why probability?

The idea that the analysis of conditionals would benefit from a probabilistic perspective has a long history. Ramsey’s original proposal can be, and often has been read as suggesting that it is what he had in mind. Later, (RT) was taken up and developed in a more explicitly probabilistic setting by Jeffrey (1964), Adams (1965), Stalnaker (1970), among others, and has since inspired a vast and varied field of inquiry.
(for recent overviews, see Edgington, 1995, and the papers in Eells and Skyrms, 1994; also Adams, 1998). The purpose of this paper is not to reiterate those points, but to propose a particular implementation of the idea which avoids certain technical problems and corrects false predictions of earlier accounts. Nevertheless, in this section I will briefly mention some reasons for exploring this direction in the first place.

2.1. ‘If...then...’ as a truth function

Minimal assumptions imply that if ‘if...then...’ is to be represented by a truth-functional connective alongside conjunction, disjunction and negation, then that connective has to be the material conditional (Gibbard, 1981; Edgington, 1986). However, the material conditional is hardly a suitable translation of ‘if...then...’, as becomes evident when we leave the realm of mathematics for which it was designed and turn to everyday uses of English. The falsehood of the antecedent is sufficient for the truth of the material conditional, but not all conditionals with false antecedents are therefore true. An assertion of (1a) is not vindicated if the match is not struck.

More generally, the material conditional is true “too easily”: It follows from premises from which one would not infer the corresponding ‘if...then...’-sentence. Instances of such inference patterns are often used to illustrate a parallel problem with counterfactual conditionals, but the same problem arises with indicatives, as the following examples show.

2.1.1. Strengthening the antecedent

\[
\frac{A \supset C}{AB \supset C}
\]

Edgington’s argument goes as follows: Let \( \mathcal{F} \) be the truth function expressed by ‘if...then...’, i.e., such that \( V(\text{if } \varphi \text{ then } \psi) = \mathcal{F}(V(\varphi), V(\psi)) \). Assume in addition that (i) sentences of the form ‘if(\varphi and \psi) then \varphi’ are tautologous and (ii) conditionals can be false. Consider two arbitrary truth-functional sentences \( p, q \). By Assumption (i), \( \mathcal{F}(V(p \text{ and } q), V(p)) \equiv 1 \). There are four cases:

<table>
<thead>
<tr>
<th>( V(p) )</th>
<th>( V(q) )</th>
<th>( V(\text{'p and q'}) )</th>
<th>( \mathcal{F}(V(\text{'p and q'}), V(p)) = 1 = \ldots )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>1</td>
<td>1</td>
<td>( \mathcal{F}(1, 1) ).</td>
</tr>
<tr>
<td>(b)</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(c)</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>(d)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Cases (a–d) exhaust three of the four possible combinations of arguments of \( \mathcal{F} \). By Assumption (ii), \( \mathcal{F}(1, 0) = 0 \).
A true material conditional cannot be rendered false by adding conditions to the antecedent. But far from contradicting (9a), (9b) is perfectly acceptable in its context. Thus the two sentences in (9) are consistent.\footnote{Examples with labels of the form \textit{nyt—}— are attested in the New York Times Corpus. The numbers refer to their locations in the corpus.}

(9)  
\begin{enumerate}
  \item If I install a better alarm system,” Griliches said, “that is an improvement in the quality of my life, and therefore a decline in inflation.
  \item But if the burglars learn how to trick this alarm system, that is a rise in price, because the quality advantage will be eroded.
\end{enumerate}

2.1.2.\textit{ Contraposition}

(10) \[
\frac{A \supset C}{C \supset A}
\]

But (11a) may be true while (11b) is false. One may assent to the former while rejecting the latter.

(11)  
\begin{enumerate}
  \item If you’re a high achiever, it takes a long time to get recognition.”
  \item If it takes a short time to get recognition, you’re not a high achiever.
\end{enumerate}

2.1.3.\textit{ Vacuous truth}

(12) \[
\frac{\overline{A}}{A \supset C}
\]

See (1) above. If the match is not lit, it does not thereby turn out that (13a) and (13b) are true, although both inferences are valid for the corresponding material conditionals.

(13)  
\begin{enumerate}
  \item If you strike the match, it will light.
  \item If you strike the match, it won’t light.
\end{enumerate}
2.1.4. Hypothetical Syllogism

(14) \[
A \supset B \\
B \supset C \\
\hline
A \supset C
\]

However, it is possible to maintain both of (15a,b), yet reject (15c).

(15) a. If I quit my job, I won’t be able to afford my apartment.
   b. If I win a Million, I’ll quit my job.
   c. ??If I win a Million, I won’t be able to afford my apartment.

Adams (1975) points out that the order of the premises matters in
this pattern. While (15) is a clear counterexample to the Hypothetical
Syllogism, this is not so obvious with (16). When the order of the
premises is changed, it is much harder to consider the case that both
are true. In linguistic terms, this is attributable to modal subordination
(Roberts, 1989), the tendency to interpret (16b) in the given context
as (16c).

(16) a. If I win a Million, I’ll quit my job.
   b. ??If I quit my job, I won’t be able to afford my apartment.
   c. ??If I win a Million and quit my job, I won’t be able to afford
      my apartment.

With the modally subordinated reading of the second premise, the
example instantiates, as a non-counterexample, a related pattern which
does seem to be valid not only with the material conditional, but also
with its natural-language counterpart:

(17) \[
A \supset B \\
AB \supset C \\
\hline
A \supset C
\]

In summary, the inference patterns listed above demonstrate that
conditionals are not faithfully represented by the material conditional.
Together with the fact that the latter is the only truth function that
could plausibly represent them, it follows that conditionals are not
truth-functional.

An alternative to the material conditional is strict implication, first
discussed by C.I. Lewis.\(^8\) According to this analysis, a conditional is

\(^8\) See Hughes and Cresswell, 1996, for a historical discussion.
true if and only if the corresponding material conditional is true at all possible worlds or, equivalently, if at all worlds at which the antecedent is true, the consequent is true. Among other advantages, it makes it possible to treat as false conditionals which under the material interpretation come out true “by accident,” for instance because their constituents happen to have the same truth value at the world of evaluation. Nor is the falsehood of the antecedent sufficient for the truth of the strict conditional.

Aside from these merits, however, the strict conditional is afflicted by many of the same problems as the material one. Contradictory antecedents still lead to vacuous truth; Strengthening of the Antecedent, Contraposition and Hypothetical Syllogism are still valid; at the same time, the strict conditional is too strong in stating in effect that the antecedent entails the consequent.

A somewhat more well-behaved variant that has taken hold in the linguistic literature on indicative conditionals is the variably strict conditional. Here the domain of the quantifier is pragmatically determined by a contextually given parameter, representing the speaker’s beliefs or some other background that the conditional is evaluated against. This contextual restriction in effect opens up a continuum of readings with the material conditional at one extreme and strict implication at the other, depending on whether the domain of the quantifier is the singleton set containing the world of evaluation, or the set of all worlds, or something in between. Thus the truth of the conditional no longer implies that the antecedent entails the consequent. However, since under this approach the conditional is true if and only if the corresponding material conditional is true at all worlds in the relevant domain, the inference patterns that were problematic for both material and strict conditional continue to be so.

2.2. Similarity

The invalidity of the inference patterns illustrated above with indicative conditionals carries over to counterfactuals; this is easy to check by substituting the corresponding forms. Counterfactuals can be, and typically are used under the assumption that the antecedent is false. Lest they be predicted vacuously true in all such contexts, the “stock of knowledge” of the preceding section is not suitable here to provide the domain of the quantifier.

Instead, in the accounts of Stalnaker (1968) and Lewis (1973), the domain is set up with the aid of a relation of similarity between worlds: (18a) is true despite the fact that there are possible worlds where the floor is covered with layers of blankets, for those worlds are less similar
to ours than those at which, as in ours, it is made of concrete. Thus it
is possible for (18a) to be true while (18b) is false.  

(18) a. If I had dropped the vase, it would have broken.
    b. If I had dropped the vase and the floor had been covered with
layers of blankets, the vase would have broken.

Stalnaker's and Lewis' theories differ in the details of the similarity
relation. Lewis maintains that for each world \( w \) and antecedent \( A \), there
may be any number (none, one, finitely or infinitely many) of "nearest"
\( A \)-worlds that are tied in their similarity to \( w \). Stalnaker maintains that
there is just one. Both agree that each world is most similar to itself,
with the consequence that at worlds at which the antecedent is true,
the truth value of a conditional is that of its consequent. In the general
case, according to Lewis a counterfactual 'If \( A \), would have been \( C \)’ is
true at \( w \) just in case there is no \( A \overline{C} \)-world that is not surpassed in
similarity to \( w \) by an \( AC \)-world. According to Stalnaker, the conditional
is true at \( w \) if and only if \( C \) is true at the nearest \( A \)-world.

While quantification over worlds is thus central to the truth of coun-
terfactuals in Lewis' theory, in Stalnaker's it enters the picture only at
the level of epistemic support.  

Let \( K \) be the set of worlds compatible with the belief state against which
the conditional is evaluated. A sentence is supported in \( K \) just in case it is true at all worlds in \( K \).

To evaluate a conditional 'if \( A \) then \( C \)' , \( K \) is transformed into a set
\( K' \) supporting \( A \) by selecting for each world in \( K \) the \( A \)-world that is
most similar to it. The conditional is supported in \( K \) if and only \( C \) is
ture at all worlds in \( K' \). 

---

9 Restricting the accessibility of worlds in this way corresponds to adding to the
antecedent a set of true sentences that are consistent with it. In this form, the idea
goes back to Goodman (1947) and is formalized in premise semantics (Kratzer,
1979, 1991a; also Veltman, 1985). For a comparison, see Lewis (1981).

10 Stalnaker (1968) assumes that the relevant set of worlds represents the beliefs
of the speaker; Stalnaker (1975) takes it to be the conversational common ground.
In either case, I prefer to speak of "support" rather than "truth," reserving the
latter for an objective, speaker-independent notion: A speaker who is wrong about
(a) may yet truthfully assert (b).

a. If you strike the match, it will light.
    b. I believe that if you strike the match, it will light.

11 Stalnaker (1975) suggests that the selection of \( A \)-worlds is restricted to those
in \( K \) whenever there are any. If there are, then \( K' \) is the set of \( A \)-worlds in \( K \): Each
\( A \)-world in \( K \) selects itself, and each non-\( A \)-world in \( K \) selects an \( A \)-world in \( K \).

If there are no \( A \)-worlds in \( K \), the selection of alternatives leads to worlds outside \( K \).

This relaxation, Stalnaker argues, is the essential semantic correlate of the difference
between indicative and counterfactual conditionals.
The similarity-based approach to regulating accessibility between worlds correctly invalidates the problematic inference patterns for counterfactuals, but not for indicatives. For the latter, Stalnaker’s account comes down to the simpler quantificational account discussed earlier.\textsuperscript{12} However, the failure to account for the inference patterns should not be held against the idea that the interpretation of conditionals involves some quantifier. It is the universal force of the quantification that re-introduces the problem. Thus the question arises whether a more suitable quantificational account can be found.

2.3. Conditional probability: The Thesis

To see how a probabilistic analysis in terms of the Thesis (T) addresses the problems pointed out above, it is helpful to consider how it is related with the quantificational account. Recall that there the truth of a conditional ‘if \( A \) then \( C \)’ relative to a set \( K \) of worlds depended on whether or not all \( A \)-worlds in \( K \) are \( C \)-worlds. The conditional is true if this is the case and false if there are any \( A \text{--C} \)-worlds in \( K \). In a probabilistic framework, the question is not whether or not, but to what extent the conditional is supported. Roughly speaking, this support is high if \( AC \) is almost as likely as \( A \), or much more likely than \( A \text{--C} \).

The first good argument for this measure comes from simple intuitions about beliefs under uncertainty. What do speakers believe about a conditional ‘if \( A \) then \( C \)’ if they are not entirely sure that \( A \text{--C} \) can be ruled out? Van Fraassen (1976) noted:

\[ \text{[T]he English statement of a conditional probability sounds exactly like that of the probability of a conditional. What is the probability that I throw a six if I throw an even number, if not the probability that: if I throw an even number, it will be a six?} \quad \text{(pp. 272–273)} \]

Most agree that the Thesis simply “sounds right,” or at least more plausible than the logical alternative. Aside from this “gut-feeling,” however, a formally more worked-out argument in favor of a probabilistic account comes from the use of conditionals in inference. Most of the relevant work in this area was carried out by Ernest Adams (1965, 1975, 1998). The two basic assumptions are paraphrased in (19).

\[ \text{(19) a. The Thesis: The probability of a conditional is the corresponding conditional probability;} \]

\textsuperscript{12} Lewis did not believe in a unified account. Instead, he followed Jackson (1979, 1987) in saying that indicatives have the truth conditions of the material conditional, combined with pragmatic “assertibility” conditions of the probabilistic kind discussed below.
b. What is preserved in everyday reasoning is not truth, but (high) probability.

Adams devised a formal system of probability-preserving inference based on (19a,b) and centered around the following notion of validity:

(20) An inference is probabilistically valid \((p\text{-valid})\) if and only if it is impossible for the premises to be highly likely while the conclusion is not highly likely.\(^{13}\)

In this system, all \(p\text{-valid}\) inferences are also classically valid, but some classically valid ones are not \(p\text{-valid}\). In particular, the following inferences with conditional conclusions are not \(p\text{-valid}\) (writing \(A \rightarrow C\) for the probabilistic conditional):

1. Strengthening the antecedent:
   \(A \rightarrow C\) does not \(p\)-entail \(AB \rightarrow C\).

2. Contraposition:
   \(A \rightarrow C\) does not \(p\)-entail \(\overline{C} \rightarrow \overline{A}\).

3. Vacuous truth:
   \(\overline{A}\) does not \(p\)-entail \(A \rightarrow C\).

4. Hypothetical syllogism (aka transitivity):
   \(A \rightarrow B\) and \(B \rightarrow C\) do not \(p\)-entail \(A \rightarrow C\).
   However: \(A \rightarrow B\) and \(AB \rightarrow C\) do jointly \(p\)-entail \(A \rightarrow C\).

A recent detailed introduction to these facts and related advantages of the theory is given in Adams (1998). I end this brief overview by concluding that a theory of conditionals in which the Thesis plays a central role holds promise in explicating the reasoning behind everyday uses of conditionals. What remains to be seen is what exactly that role should be.

2.4. INTERIM SUMMARY

This ends the short overview of the relevant background. Before moving on, I will briefly what follows in the perspective of this section and add some historical remarks.

\(^{13}\) The formal definition is as follows: A sentence \(\varphi\) follows from a set of sentences \(\Gamma\) if and only for all \(\epsilon > 0\) there is a \(\delta > 0\) such that for all conditional probability functions \(Pr\) such that \(Pr(B) \geq 1 - \delta\) for all \(B \in \Gamma\), \(Pr(\varphi) \geq 1 - \epsilon\).
Two recurring themes in quantificational accounts of conditionals are the force of the quantification and its domain. Between the two main ingredients of this paper, the labor is divided accordingly: The use of probabilities addresses the former, the appeal to causal relations the latter.

The case for adopting a probabilistic perspective was outlined above. There are certain choices to be made and technical problems to be overcome in integrating a probabilistic component in the familiar possible-worlds semantics. These issues will be addressed in Sections 3 and 4.

I will then turn to the reasons for making the interpretation sensitive to causal relations in Section 5. From the perspective of the logical accounts discussed in this section, this solution is formally closest to the set selection approach, a special case of the Stalnaker/Lewis similarity-based framework (Lewis, 1973 attributed the idea to John Vickers and Peter Woodruff) which arises under the assumption that there always is a non-empty, but not necessarily singleton set of closest antecedent worlds. The use I propose to make of causal relations can be thought of as defining the analog of a set selection function that is not based on an unanalyzed relation of similarity.

Similarity between worlds has also been put to use in probabilistic accounts, where it gives rise to the update operation of imaging, as opposed to conditionalization (Stalnaker, 1970; Lewis, 1976; Gärdenfors, 1982, 1988). Skyrms (1984, 1988, 1994) made extensive use of the analog of set selection in a probabilistic setting, also referring to causality for motivation. Skyrms' account thus is closest to the one I am going to propose. Despite this similarity, however, there are several important differences in the use of causal relations, details of the implementation, and the predictions about the probabilities of conditionals, especially counterfactuals. These differences deserve a detailed comparison, but although I will briefly return to them at the end, I leave a full discussion for another occasion.

3. Implementation

The previous section made a case for a probabilistic analysis of conditionals. I will now say more on the notions of probability and truth, and the relationship between them.

3.1. Probabilities of propositions

Uncertainty about the world is modeled as usual, using sets of possible worlds to represent open, "live" possibilities. The quantifiers of
modal logic allow us to express statements such as whether something is necessary, possible, or impossible with respect to such a set of worlds. With probabilities, we can in addition assign intermediate degrees of “support” to propositions. Starting with \( W \), the set of all worlds, a probability distribution is defined as follows.

**Definition 1 (Probability model)**

A probability model is a structure \( \langle W, Pr \rangle \), where \( W \) is a non-empty set of worlds and \( Pr \) is a probability distribution over \( W \), i.e., a function \( Pr : \wp(W) \rightarrow [0, 1] \) such that for all \( X, Y \subseteq W \):

\[
Pr(W) = 1 \\
0 \leq Pr(X) \leq 1 \\
Pr(X \cup Y) = Pr(X) + Pr(Y) \\
\text{if } X \text{ and } Y \text{ are disjoint}\)

The conditional probability of \( Y \), given \( X \), is defined as the “amount” of \( Y \)-worlds within the set of \( X \)-worlds, whenever the latter has positive probability.

**Definition 2 (Conditional probability)**

The conditional probability of \( Y \), given \( X \) for \( X, Y \subseteq W \) is defined as follows:

\[
Pr(Y|X) = \begin{cases} 
\frac{Pr(X \cap Y)}{Pr(X)} & \text{if } Pr(X) \neq 0 \\
\text{undefined otherwise} & 
\end{cases}
\]

The conditional probability is undefined if \( X \) has zero probability. This is above all an artifact of the mathematical definition; one may make use of or dispose of it in various ways. One way is to stipulate some value for the conditional probability in those cases, such as 1 (Adams, 1965; McGee, 1989), \( Pr(Y) \) (Milne, 1997), or any arbitrary value (Skyrms, 1988).

Alternatively, with an eye toward the application to conditionals, we may consider the undefined value a virtue rather than a shortcoming. Linguistic theories often incorporate the observation that there is nothing “correct” to be believed about a conditional when the antecedent is definitely false by stipulating that conditionals presuppose that there

---

\(^{14}\) The third condition must hold in general for the limits of countable unions of pairwise disjoint propositions. This is important for the coherence of the measure \( Pr \), but I will ignore it here for simplicity.
are antecedent worlds to be quantified over. Assuming that presupposition failure results in lack of a truth value, this stipulation may be seen as the logical analog to the probabilistic fact that the probability of a conditional is undefined if its antecedent has zero probability.

3.2. Probabilities of truth-functional sentences

The last section defined probabilities of propositions. The probabilities of sentences are defined in terms of those.

Definition 3 (Language $L_A^1$)
Given a set $A$ of propositional letters, the language $L_A^1$ is the smallest set containing $A$ and closed as follows: If $\varphi, \psi \in L_A^1$, then $\overline{\varphi}, \varphi \psi \in L_A^1$.

Note that $\overline{\varphi \psi}$ is the negation of $\varphi \psi$, whereas $\overline{\varphi}$ is the conjunction of $\varphi$ and $\overline{\psi}$. Disjunction and material conditional may be defined as usual: $\varphi \lor \psi$ and $\varphi \supset \psi$ are $\overline{\varphi \overline{\psi}}$ and $\varphi \overline{\psi}$, respectively. I refer to the material conditional using the symbol $\supset$. We already know, of course, that the material conditional is not a suitable connective to model the natural-language conditional. Below, I will use $\rightarrow$ for that “natural” conditional.

The truth values 1 and 0 are assigned to sentences pointwise at individual worlds by an interpretation function $V^1$.

Definition 4 (Interpretation for $L_A^1$)
An interpretation of the language $L_A^1$ in a probability model $(W, Pr)$ is a function $V^1 : L_A^1 \rightarrow \{0, 1\}^W$ satisfying the following conditions:

For $A \in A : V^1(A)(w) \in \{0, 1\}$

$V^1(\overline{\varphi})(w) = 1 - V^1(\varphi)(w)$

$V^1(\varphi \psi)(w) = V^1(\varphi)(w) \cdot V^1(\psi)(w)$

For each $\varphi$, $V^1(\varphi)$ is the characteristic function of the set of worlds at which $\varphi$ is true. In a probabilistic context, functions from possible worlds to truth values are a special case of random variables.\(^{15}\) By convention, the range of a random variable is the set of real numbers or a subset thereof, such as $\{0, 1\}$ or $[0, 1]$. Random variables may in general be continuous. I will ignore that case throughout for the sake

\(^{15}\) In statistical jargon, the set of possible worlds is the sample space, in which each world is an outcome. A proposition is an event; its characteristic function is an indicator function. The statistical notion of an event is not to be confused with the “events” often dealt with in natural-language semantics. Instead, it is best thought of as the “event” of the English phrase ‘in the event that . . . ’
of simplicity. To generalize to the continuous case, summations are substituted with integrals.

As we will see, there is a certain tension between the probabilistic account and the idea that conditionals denote (characteristic functions of) propositions, as other sentences do. I will not make this latter assumption: The denotations of conditionals will be random variables, but not ones whose range can in general be restricted to \{0, 1\}.

I will have occasion to refer to the set of worlds at which a sentence takes a particular value; for instance, the set \( \{ w \in W | V^1(A)(w) = 1 \} \) is the proposition “that \( A \) is true.” In the interest of readability, I abbreviate this expression as \( V^1(A) = 1 \); thus \( Pr( \{ w \in W | V^1(A)(w) = 1 \} ) \) is alternatively written \( Pr(V^1(A) = 1) \). Furthermore, I will write \( Pr(V^1(A) = 1, V^1(B) = 1) \) instead of \( Pr( \{ w \in W | V^1(A) = 1 \} \cap \{ w \in W | V^1(B) = 1 \} ) \) for the proposition that both \( A \) and \( B \) are true.

There is a straightforward relationship between the values of the characteristic function of a proposition and that proposition’s probability: The latter equals the expectation of the former—i.e., the weighted sum of its values, where the weights are the probabilities that the function takes those values. This notion, as well as the more general one of conditional expectation, will be important below.

**Definition 5 (Expectation)**

For random variables \( X, Y \), expectation and conditional expectation are defined as follows:

\[
E[X] = \sum_{x \in \text{range}(X)} x \cdot Pr(X = x)
\]

\[
E[X|Y = y] = \sum_{x \in \text{range}(X)} x \cdot Pr(X = x|Y = y)
\]

Here \( X \) stands for any random variable, such as the denotation \( V^1(\varphi) \) of a sentence \( \varphi \). The probability of a sentence is defined as the expectation of the characteristic function it denotes.

**Definition 6 (Probabilities of sentences)**

Given a probability model \( \langle W, Pr \rangle \) and an assignment \( V^1 \) for \( \mathcal{L}_A^1 \), a probability distribution \( P \) on the sentences in \( \mathcal{L}_A^1 \) is defined as follows: For all \( \varphi, \psi \in \mathcal{L}_A^1 \),

\[
P(\varphi) = E[V^1(\varphi)]
\]

\[
P(\psi|\varphi) = E[V^1(\psi)|V^1(\varphi) = 1]
\]
Since all sentences in $L^1_A$ denote functions with range $\{0, 1\}$, the summation in Definition 5 is trivial and the relationship between $P$ and $Pr$ is straightforward:

**Fact 1**

For all $\varphi, \psi$ in $L^1_A$:

a. $P(\varphi)$ is the probability that $\varphi$ is true:

\[ P(\varphi) = E[V^1(\varphi)] = \sum_{x \in \{0, 1\}} x \cdot Pr(V^1(\varphi) = x) \]
\[ = 0 \cdot Pr(V^1(\varphi) = 0) + 1 \cdot Pr(V^1(\varphi) = 1) \]
\[ = Pr(V^1(\varphi) = 1) \]

b. $P(\psi|\varphi)$ is the conditional probability that $\psi$ is true, given that $\varphi$ is true:

\[ P(\psi|\varphi) = E[V^1(\psi)|V^1(\varphi) = 1] \]
\[ = \sum_{x \in \{0, 1\}} x \cdot Pr(V^1(\psi) = x|V^1(\varphi) = 1) \]
\[ = 0 \cdot Pr(V^1(\psi) = 0|V^1(\varphi) = 1) + 1 \cdot Pr(V^1(\psi) = 1|V^1(\varphi) = 1) \]
\[ = Pr(V^1(\psi) = 1|V^1(\varphi) = 1) \]

With these definitions, a number of facts about the probabilities of sentences follow as expected (given without proof):

**Fact 2**

For all $\varphi, \psi \in L^1_A$:

\[ P(\varphi) = 1 \text{ if } \varphi \text{ is a tautology} \]
\[ 0 \leq P(\varphi) \leq 1 \]
\[ P(\varphi \lor \psi) = P(\varphi) + P(\psi) \text{ if } P(\varphi \psi) = 0 \]
\[ P(\varphi) \leq P(\psi) \text{ if } \varphi \text{ entails } \psi \]
\[ P(\overline{\varphi}) = 1 - P(\varphi) \]
\[ P(\varphi \psi) = P(\varphi) \cdot P(\psi|\varphi) \]
\[ P(\varphi \supset \psi) = 1 - (P(\varphi) - P(\varphi \psi)) \]

3.3. Probabilities of conditionals

The probability of the sentence ‘you strike the match’ is the probability that you strike the match. The Thesis asserts that the probabilities
of conditionals are conditional probabilities. The goal is to extend the truth value assignment $V^2$ to conditionals in such a way that the expectation $P(A \rightarrow C)$ of the truth values of $A \rightarrow C$ equals the conditional probability $P(C|A)$ whenever the latter is defined, regardless of the underlying probability distribution $Pr$.

Unfortunately, it is impossible, aside from certain uninteresting special cases, to assign truth values to conditionals in this way. A conditional probability is not the probability that a proposition is true. Consequently, if conditionals are to obey the Thesis, they cannot denote propositions like atomic and truth-functional sentences. This is the lesson from Lewis’ (1976, 1986b) triviality results and a large body of subsequent work.

3.3.1. Triviality
Lewis’ results and their various extensions and generalizations are introduced in a number of detailed discussions (Gibbard, 1981; Hájek and Hall, 1994; Hájek, 1994; Edgington, 1995, and elsewhere). I will only give an informal description of one of them in the framework introduced in Section 3. This will illustrate an essential part of the problem, and it will help the reader to see how the particular solution I am going to adopt avoids it.

The assumption that conditionals denote propositions in the usual sense implies that their values are, at each individual world, (i) either 0 or 1 and (ii) independent of the probability distribution. In particular, $A \rightarrow C$ will retain its interpretation under conditioning, so we may ask what $P(A \rightarrow C|D)$ should be when $P(D)$ is not zero. Lewis assumes that this probability should equal $P(C|AD)$ whenever the latter is defined. Consider the special case in which $D$ is either $C$ or $\overline{C}$. Then the assumption implies that

$$
(21a) \quad P(A \rightarrow C | C) = 1
$$

$$
(21b) \quad P(A \rightarrow C | \overline{C}) = 0
$$

This is plausible. In case the coin is rigged to come up tails no matter what, the conditional in (22a) is certainly true and (22b) is certainly false.

(22) a. If I bet on tails, the coin will come up tails.
    b. If I bet on tails, the coin will come up heads.

To see the formal consequences of this assumption, recall that conditional probabilities of sentences equal the underlying conditional expectations. So (21a,b) amount to (23a,b):
(23a) \[ E[V(A \rightarrow C)|V(C) = 1] = 1 \]
(23b) \[ E[V(A \rightarrow C)|V(C) = 0] = 0 \]

Thus with probability 1, the conditional is equivalent to its consequent. Recall also that the above should hold for any underlying probability distribution \( Pr \) in which the conditional probabilities are defined. As a consequence, (24a) must be the case for any corresponding distribution \( P \) over sentences. Together with the Thesis, this implies (24b).

(24a) \[ P(A \rightarrow C) = P(C) \]
(24b) \[ P(C|A) = P(C) \]

It follows that the consequent must be stochastically independent of the antecedent. This is absurd: Conditionals are typically used to assert that such a dependence does hold. In addition, it implies that the Thesis only holds in full generality—i.e., for all distributions in which the probabilities are defined—if the language distinguishes at most two non-empty propositions, for (24a) is inconsistent with the Thesis whenever three mutually disjoint propositions \( AC, \overline{AC} \) and \( \overline{A} \) have non-zero probability.\(^{16}\) Lewis calls such a language trivial.

This first result shows that unless the language is trivial, its connective ‘\( \rightarrow \)’ cannot be a “universal probability conditional,” that is, its interpretation cannot be independent of the probability distribution. It is possible, however, to start out with a given probability distribution and define a conditional in such a way that the Thesis holds within that particular distribution. In his remaining triviality results, Lewis considers what happens if such a conditional that is “tailored” to a given probability distribution retains its interpretation while the probabilities are altered by various kinds of update operations. He shows that except for special trivial cases, it is impossible to guarantee that the Thesis holds both before and after the update. Lewis’ results have since been extended and generalized in a number of ways (see the references mentioned above).\(^{17}\)

---

\(^{16}\) Consider \( A \rightarrow AC \). (i) By (24a), \( P(A \rightarrow AC) = P(AC) \). (ii) By the Thesis, the probabilistic calculus and propositional logic, \( P(A \rightarrow AC) = P(AC|A) = P(ACA)/P(A) = P(AC)/P(A) \). Hence by (i) and (ii), \( P(A) = 1 \) and \( P(\overline{A}) = 0 \).

\(^{17}\) Aside from the probabilistic setting with which these elaborations are mostly concerned, an interesting related result was given by Gärdenfors (1988) for the non-probabilistic case.

---

kaufmann@northwestern.edu
Draft - do not cite. Comments welcome.
July 6, 2003
3.3.2. Context dependence
The triviality results show that the Thesis is incompatible with the assumption that conditionals denote propositions in the usual sense. It does not follow that the Thesis cannot be upheld. One way to avoid triviality is to assign values to the conditional that in turn depend on the probability distribution, thus making its denotation dependent upon the context in which it is interpreted.

Various authors have found good arguments for such a position, both in probabilistic and non-probabilistic settings (Harper, 1976; Stalnaker, 1975; Gibbard, 1981; van Rooy, 1997). How best to make sense of it depends on the relevant notion of “context.” Some of the mentioned authors explicitly made their arguments with regard to epistemic conditionals and subjective belief states. I adopt it here for predictive conditionals and objective chance. The question of what constitutes the context in this case and in what sense it is objective will be taken up in Section 4.

Specifically, I will adopt, but modify below, a proposal made by Jeffrey (1991; cf. also Stalnaker and Jeffrey, 1994). Jeffrey, in a sense, put the cart in front of the horse: Instead of deriving the probability of the conditional from considerations of what its truth values at individual worlds should be, he started out with a given amount of probability mass—the conditional probability—and explored a systematic way of distributing it over the set of possible worlds. His proposal rests on the following simple fact about the probabilistic calculus:

\begin{align}
P(\psi|\varphi) &= P(\psi|\varphi)[P(\varphi) + P(\overline{\varphi})] \\
&= P(\varphi\psi) + P(\psi|\varphi)P(\overline{\varphi}) \\
&= 0 \cdot P(\varphi\overline{\psi}) + 1 \cdot P(\varphi\psi) + P(\psi|\varphi) \cdot P(\overline{\varphi})
\end{align}

Each of the terms on the right-hand side of (26) may be read as the product of the value of a random variable and the probability with which it takes that value. The sentences determining the weights are mutually incompatible and jointly exhaust all possibilities. Conditionals, Jeffrey suggests, denote random variables defined in just this way. Recalling that \( P(\psi|\varphi) = E[V(\psi)|V(\varphi) = 1] \), the corresponding value assignment may be defined as in (27).

\begin{align}
V(\varphi \rightarrow \psi)(w) &= \begin{cases} 
V(\psi)(w) & \text{if } V(\varphi)(w) = 1 \\
E[V(\psi)|V(\varphi) = 1] & \text{if } V(\varphi)(w) = 0
\end{cases}
\end{align}
This is, of course, only one among countless ways of distributing values over worlds in such a way that their expectation is the conditional probability. It is particularly appealing, however, in that it assigns to the conditional a definite truth value (that of the material conditional) at worlds at which the antecedent is true, in accordance with the intuition that in that case the truth or falsehood of the conditional can be determined by simple inspection of the facts. Where the antecedent is false, the conditional takes values that may lie between 0 and 1 and depend, via the expectation, on the probability distribution.

Thus the assignment in (27) ensures by its very definition that the Thesis will be upheld. It is not immediately clear, however, in what sense this “rather weird three-valued entity” (Edgington, 1995) expresses what conditionals mean. Jeffrey, whose concerns are purely “top-down” (starting out with the desired probability and distributing values accordingly), does not offer much in the way of an intuitive motivation, but other authors have given such a motivation and argued that it can in fact be gleaned from the set-selection semantics in the tradition of Stalnaker and Lewis (van Fraassen, 1976; Stalnaker and Jeffrey, 1994; Edgington, 1995). Recall that in Stalnaker’s theory, the value of $A \rightarrow C$ at worlds at which $A$ is false depends on the nearest $A$-world, which is uniquely identified in the model by a selection function. (27) does not presuppose the existence of a single nearest $A$-world; instead, the selection function picks an $A$-world at random. The conditional is guaranteed by Definition (27) to have a definite truth value there. The expectation of that truth value is just the conditional probability of the consequent, given the antecedent. This is the value assigned to the conditional at the non-antecedent world of evaluation.

I will use Jeffrey’s definition (27) as a point of departure in developing my account. In the next section, I will say some more on the justification of the assignment function in (27). In particular, two questions need to be addressed: First, as defined above, $V$ is a function of three arguments: A sentence, a world, and a probability distribution. The value at one and the same world will change if the probability distribution changes. Is there a sense, then, in which that value can be thought of as “objective,” like truth values? Secondly, what does

\[ P(\varphi \rightarrow \psi) = E[V(\varphi \rightarrow \psi)] = \sum_{x \in [0, 1]} x \cdot Pr(V(\varphi \rightarrow \psi) = x) = 0 \cdot Pr(V(\varphi \rightarrow \psi) = 0) + 1 \cdot Pr(V(\varphi \rightarrow \psi) = 1) + E[V(\psi)|V(\varphi) = 1] \cdot Pr(V(\varphi \rightarrow \psi) = E[V(\psi)|V(\varphi) = 1]|V(\varphi) = 1) = 0 \cdot Pr(V(\varphi) = 0, V(\psi) = 0) + 1 \cdot Pr(V(\varphi) = 1, V(\psi) = 1) + E[V(\psi)|V(\varphi) = 1] \cdot Pr(V(\varphi) = 1) = 0 \cdot P(\varphi \psi) + 1 \cdot P(\varphi \psi) + P(\psi|\varphi)P(\varphi) = P(\varphi) + P(\psi) - P(\varphi)P(\varphi) = P(\psi|\varphi) \]
it mean for these values at individual worlds—in the absence of any uncertainty about the facts—to be intermediate?

4. Truth and chance in time

Predictive conditionals are used to make conditional predictions. Their formal treatment must involve some representation of time. In this section I will extend the model accordingly.

First of all, however, some preliminary comments are in order. Specifically, talk of “truth” must be made precise when it comes to future reference. Since I defined the probabilities of sentences as the expectations of their truth values, in order to speak of the (present) probabilities of predictions, it must be ensured that they already have truth values. Some speakers find this counterintuitive and maintain that except in cases of predetermination, sentences about the future are generally “not yet true” (or false) at the time they are used. This judgment is fully legitimate, but the notion of truth that it appeals to is not the “Ockhamist” one that I have in mind.

It is well-known that the past is settled in a sense in which the future is not. This asymmetry has been discussed at least since Aristotle’s remarks on future sea battles (On Interpretation 1:9), and more recently by Ramsey (1929), Reichenbach (1956), Prior (1967), Thomason (1970), Burgess (1979), Lewis (1979, 1980), and others. The difference is that at their respective evaluation times, (28a) is either necessarily true or necessarily false, whereas (28b) is neither necessarily true nor necessarily false.

(28) a. The coin landed heads.
   b. The coin will land heads.

The central tenet of the Ockhamist view is that (28b) does have the weaker property of being necessarily either true or false: It is true at the time it is used if and only if it turns out true at the relevant later time. Accordingly, it may be true without there being any way of knowing (already) that it is true.

On the alternative, “Peircean” view, there is no use for truth values that exist but elude us when they are needed most (namely when the sentence is uttered); instead, for the sentence to be true, its truth must turn out true in all possible continuations of history. Most predictions therefore fail to have truth values.

Thus the Peircean notion of “truth” coincides with Ockhamist “settledness.” This is the notion behind the intuition that sentences like (28b)
have no truth values at utterance time. In this paper, in contrast, the truth values whose expectation is the probability of a prediction are of the Ockhamist kind.

4.1. Models

To incorporate time into the model, the basic structure I adopt is what Thomason (1984) calls a \( T \times W \)-frame.\(^{19} \) Worlds, structureless points in Section 3, are now “stretched” along the temporal dimension and represented as complete trajectories. Where necessary to avoid confusion, I will refer to them as “world-lines.” A history is a decreasing sequence of sets of such trajectories, recording at each time \( t \) the past facts accumulated up to \( t \) as well as the continuations possible at \( t \). Technically, this is embodied in the condition that all worlds included in the time slice of the history at \( t \) be indistinguishable at all times up to \( t \) but may come apart thereafter.

Figure 1 visualizes the evolution of such a history. Three histories are represented by narrowing gray areas, each including ever fewer worlds as time progresses (from left to right). The black areas enclose the

\(^{19} \) Some of the definitions in this section are adapted from Thomason (1984), where the reader is also referred for a discussion of the relationship between \( T \times W \)-frames and models of branching time.
worlds that agree in all facts up to some particular time. These are each
other’s historical alternatives at that time. It is with respect to this set
of alternatives that time-variant notions which refer to quantities of
worlds, such as objective chance, are encoded.

The definition of $T \times W$-frames is adopted from Thomason (1984).
The intention is that $T$ is a set of moments in time, linearly ordered
by the earlier than relation $<$, and $W$ is the set of worlds.

**Definition 7** ($T \times W$-frames (Thomason, 1984))

A $T \times W$-frame is a quadruple $\langle W, T, <, \approx \rangle$, where

1. $W$ and $T$ are disjoint nonempty sets,
2. $<$ is a transitive relation on $T$ which is also
   (a) irreflexive: for all $t \in T$, $t \neq t$; and
   (b) linear: for all $t, t' \in T$, either $t < t'$ or $t' < t$ or
       $t = t'$;
3. $\approx$ is a three-place relation in $T \times W \times W$, such that
   (a) for all $t, \approx_t$ is an equivalence relation; and
   (b) for all $w, w' \in W$ and $t, t' \in T$, if $w \approx_t w'$ and $t' < t$,
       then $w \approx_t w'$.

Each world evolves by shedding alternative futures. I will refer to the
equivalence class $\{w' \ | \ w \approx_t w'\}$ of historical alternatives of world $w$ at
time $t$ as $[w]_t^\approx$. Clause (3b) in the definition is intended to ensure that
the progressive loss of alternatives is not reversible. It is not possible for worlds to “become” historical alternatives at some point in time, or
for alternatives “not to have been” alternatives at earlier times.

The assignment function $V$ is now defined in such a way that each
atomic sentence at each time is assigned (the characteristic function of)
the set of worlds in which it is true at that time.

**Definition 8** ($T \times W$ interpretation)

An interpretation of the language $\mathcal{L}_A$ in a $T \times W$-frame is a function $V :$
$\mathcal{L}_A \mapsto (T \mapsto (W \mapsto \{0, 1\}))$ from expressions in $\mathcal{L}_A$ to functions from
members of $T$ to (characteristic functions of) propositions, provided that for all atomic sentences $A$ in $\mathcal{A}$ and all times $t$, if $w \approx_t w'$, then
$V(A)(t)(w) = V(A)(t)(w')$.

Propositions, sets of worlds as before, are now sets of world-lines.
The added temporal dimension affords a richer set of means by which
to identify them, using sentences like those in (29).

(29) a. John has been to New York.
    b. John was in New York at 5am on July 2nd, 1999.
    c. John was in New York at 5am today.
d. John was in New York.

The sentence in (29a) involves existential quantification over times, (29b) contains the proper name of a day, (29c) a function returning ‘5am that day’ for each time of evaluation, and (29d), when used felicitously, co-refers with a contextually given reference time. For the purposes of this paper, I will not be concerned with any of these distinctions, assuming only that sentences are somehow given their reference times. Formally, I superscript sentences with elements of T, the times in the model. A cleaner way would be to use a set of distinct constants which are interpreted as referring to those times, but no confusion is likely to arise from the shortcut.

**Definition 9 (Language \( \mathcal{L}_A^{T_1} \))**

The language \( \mathcal{L}_A^{T_1} \) is the smallest set containing \( A \) and such that for all \( \varphi, \psi \in \mathcal{L}_A^{T_1} \) and \( t \in T \), \( \varphi, \varphi^t, \varphi^d \in \mathcal{L}_A^{T_1} \).

The assignment of truth values is now as given in Definition 10.

**Definition 10 (Interpretation for \( \mathcal{L}_A^{T_1} \))**

An interpretation \( V \) for the language \( \mathcal{L}_A \) in a model \( \langle W, T, <, \sim \rangle \) (cf. Definition 8) is extended to an assignment \( V^T \) for \( \mathcal{L}_A^{T_1} \) as follows: For all \( \varphi \in \mathcal{L}_A^{T_1} \),

\[
\text{If } \varphi \in \mathcal{L}_A : V^T(\varphi)(t)(w) = V(\varphi)(t)(w)
\]

\[
V^T(\varphi^t)(t)(w) = V^T(\varphi)(t^t)(w)
\]

I will henceforth drop the superscripts from \( V^T \) and \( \mathcal{L}_A^{T_1} \).

“Chance” is the objective probability that a proposition is true, following (Lewis, 1980; cf. also van Fraassen, 1981). It is encoded as a prior probability distribution over the worlds (now world-lines) in the model, similarly to the case of the simpler models of Definition 1.

**Definition 11 (Chance model)**

A chance model is a structure \( \langle W, T, <, \sim, Pr \rangle \) where \( \langle W, T, <, \sim \rangle \) is a \( T \times W \)-frame (cf. Definition 7) and \( Pr \) is a probability distribution over \( W \).

Chance does not remain constant over time. In the model, the development of the probability distribution through history proceeds by conditionalization on what is settled. What is settled in a world \( w \) at a time \( t \) is represented by the proposition \( [w]_t^{\sim} \).

**Definition 12 (Chance in history)**

The chance of a proposition \( X \) at a time \( t \) in a history \( h \) is \( Pr(X|[w]_t^{\sim}) \).
Thus “interesting” chances arise only with respect to the future. The chance that the match was struck is either 1 (if it was struck) or 0 (if it wasn’t). The chance that it will be struck may fall between 1 and 0.

What is important about the above definitions is that chances are fully determined by the world and time of evaluation, since these two parameters determine the set of historical alternatives in Definition 12. In this sense, the values of conditionals, though dependent upon the probability distribution, are nevertheless objective. Thus a potential objection to the Jeffrey conditional is avoided which already voiced by (Lewis, 1976), who insisted that the conditional must have an interpretation that is independent of subjective probabilities: “Else how are disagreements about a conditional possible, or changes of mind?” Assuming an objective interpretation of the probability, Lewis’ question concerns the relationship between it and subjective probabilities, which is beyond the scope of this paper.20

4.2. Interpretation

The probabilities of sentences, too, depend on the time of evaluation. In addition, they depend on the world, since different histories give rise to different probability distributions over the continuations. We can therefore write:21

\[
(30) \ P_{w,t}(\varphi) = E[V(\varphi)(t)|[w]_t^\infty] = \sum_{x \in [0,1]} x \cdot Pr(V(\varphi)(t) = x|[w]_t^\infty) = Pr(V(\varphi)(t) = 1|[w]_t^\infty) \text{ if } \varphi \text{ is truth-functional}
\]

Conditionals are evaluated according to Jeffrey’s proposal in (27) on page 22, with the consequence that their probabilities equal the conditional probabilities at the time of evaluation:

---

20 In defining chance as a prior probability distribution which evolves by conditioning, I sidestepped some deep metaphysical questions, such as how chance is related to the more “mundane” facts of the world, whether it could in principle be known (Lewis, 1980, 1994) and, indeed, whether there it exists. I will not take a stance on these questions. I take it to be self-evident that speakers talk as if there was such a thing, but for my purposes it may just as well be a special kind of subjective belief—special in that we take it to be the beliefs that anyone would hold who knew everything about the past and could correctly project expectations from the past.

21 Recall that the right-hand side of the first line is can be further rewritten as \( \sum_{x \in [0,1]} x \cdot Pr([w|w \approx_t w', V(\varphi)(t)(w') = x]) / Pr([w]_t^\infty) \).
\[ V(\varphi \rightarrow \psi)(t)(w) = \begin{cases} V(\psi)(t)(w) & \text{if } V(\varphi)(t)(w) = 1 \\ E[V(\psi)(t)|V(\varphi)(t) = 1, [w]_t] & \text{if } V(\varphi)(t)(w) = 0 \end{cases} \]

I will drop the subscripts $w, t$ later on. The connection between the values and the probability of a conditional is again straightforward (the argument is similar to Footnote 18).

\[ P_{w,t}(\varphi \rightarrow \psi) = E[V(\varphi \rightarrow \psi)(t)|[w]_t] \]
\[ = Pr(V(\psi)(t) = 1|V(\varphi)(t) = 1, [w]_t) \]

As before at worlds at which the antecedent is true, the conditional has the truth value of its consequent. Also as before, at worlds at which the antecedent is false, the conditional may receive an intermediate value.

It is not hard to find a conceptually satisfactory explanation of this consequence in the context of time and chance. According to the Ockhamist view, the current truth or falsehood of a non-conditional prediction at a given world is determined by future facts. The uncertainty arises because it is impossible to tell which of a number of alternative worlds is the actual one, not because the actual world is such that the prediction does not yet have a truth value. Taken as a guide in discovering the truth value, this notion implies that once the relevant facts are settled, one should be able to tell in retrospect whether the prediction was true or false at the time it was made. Such a finding would not imply that it was likely to be true at the time, hence that the speaker was justified in making it—it may have turned out true by pure luck.

The same is only partly the case for conditional predictions like (33). As it happens, both antecedent and consequent are true, so the sentence was true throughout the decades leading up to the events in question. This does not imply, however, that its truth was foreseeable.\(^{22}\)

(33) If Gorbachev visits East Berlin in October, 1989, the Wall will fall in November, 1989.

But what will the future reveal about the truth of a conditional whose antecedent turns out false? I agree with Ramsey (1929), who

---

\(^{22}\) Mackie (1973) discussed conditionals that turn out true or false by coincidence (see also Edgington, 1995). Pendlebury (1989) and Read (1993) hold the opposing view that the conjunction should not entail the conditional.
maintains that the truth or falsehood of such “unfulfilled” conditionals is not determined by the facts. In a world in which the antecedent of (33) is false, there seems to be no sure way to determine, at any subsequent time, whether the conditional was true or false.

For a less “historical” and perhaps even more convincing example, consider an upcoming toss of a fair die. There surely is a number $x$ such that the outcome will be $x$. It is also clear that if the outcome is an even number, then there will be a number $y$ such that the outcome is $y$. But unless the outcome is in fact an even number, there is no number $z$ such that if the outcome is an even number, it will be $z$.

But even if the antecedent turns out false, the world did “support” the conditional to some extent, given the facts up to the time in question. This may be taken as an intuitive rationale for assigning some value in this case, rather than leaving it undefined.

5. Right-nested conditionals

The previous sections showed how the random-variable interpretation of conditionals circumvents the formal problem highlighted by the triviality results and how it may be integrated in an overall probabilistic approach to sentence interpretation. One of the great merits of this approach is that it yields predictions about the probabilities of compounded and embedded conditionals, such as (34a), which I consider equivalent to the variant in (34b).

\[ (34) \quad \begin{array}{ll} 
  a. & \text{If the match is wet, then if you strike it, it will light.} \\
  b. & \text{If the match is wet, then it will light if you strike it.} \\
  c. & W \rightarrow (S \rightarrow L) 
\end{array} \]

The standard probabilistic calculus does not provide probabilities of expressions of the form (34c): `\( P((L|S)|W) \)` is not defined. But it is, of course, possible to take as the probability of (34c) the expectation of the values of consequent $S \rightarrow L$ over those worlds at which the antecedent $W$ is true, and assign that expectation value at those worlds at which $W$ is false. For right-nested conditionals, no modification of Jeffrey’s truth assignment is needed to accomplish this.23

Unfortunately, however, the predictions that Jeffrey’s account makes for such conditionals turn out to be counterintuitive (Edgington, 1991; Lance, 1991). To see this, consider a concrete example.

---

23 The interpretation of left-nested and conjoined conditionals involves some complications; see Stalnaker and Jeffrey, 1994, for these cases, which I do not discuss in this paper.
(35) The probability that...

a. it gets wet is “low” (.1)

b. you strike it is .5

c. it lights given that you strike it and it is dry is “high” (.9)

d. it lights given that you strike it and it is wet is “low” (.1)

and the probabilities of the striking and the wetness are independent of each other.

Consider how in this scenario the values of the consequent (36) are distributed over the set of worlds.

(36) If you strike the match, it will light.

In Figure 2 and below, the values are represented as shades of gray: black for the value 1 at worlds at which the match is struck and lights, white for the value 0 at worlds where it is struck and does not light, and gray for the intermediate value .82 assigned at those worlds at which it is not struck. The value assignment is as in (37) (for the sake of readability, I omit reference to times here and later on in the section).

---

24 The value .82 is the conditional probability that it lights, given that it is struck and can be calculated from the numbers given above as follows (recall that W and S are independent, so \(P(W|S) = P(W)\) and \(P(W|\bar{S}) = P(\bar{W})\)):

\[
P(L|S) = \frac{P(LS)P(S)}{P(S)} = \frac{[P(LSW) + P(LS\bar{W})]P(S)}{P(S)}
\]

\[
= \frac{P(L|SW)P(W|S)P(S) + P(L|S\bar{W})P(\bar{W}|S)P(S)}{P(S)}
\]

\[
= P(L|SW)P(W) + P(L|S\bar{W})P(\bar{W}) = .1 \cdot .1 + .9 \cdot .9 = .82
\]
Figure 3. Distribution of values of (34)

\[
V(S \rightarrow L)(w) = \begin{cases} 
V(L)(w) & \text{if } V(S)(w) = 1 \\
E[V(L)|V(S) = 1] & \text{if } V(S)(w) = 0 \\
0 & \text{if } V(S)(w) = 1, V(L)(w) = 0 \\
1 & \text{if } V(S)(w) = 1, V(L)(w) = 1 \\
.82 & \text{if } V(S)(w) = 0 
\end{cases}
\]

The conditional in (34) is evaluated as follows: The values of its consequent (36) are assigned at those worlds at which the antecedent is true, i.e., at which the match is wet. The expectation of these values is then distributed over the set of worlds at which the antecedent is false, i.e., the match is dry. This expectation is calculated in (38).

\[
E[V(S \rightarrow L)|V(W) = 1] = \sum_{x \in [0,1]} x \cdot Pr(V(S \rightarrow L) = x|V(W) = 1) \\
= 0 \cdot Pr(V(S) = 1, V(L) = 0|V(W) = 1) \\
+ 1 \cdot Pr(V(S) = 1, V(L) = 1|V(W) = 1) \\
+.82 \cdot Pr(V(S) = 0|V(W) = 1) \\
= 0 \cdot .45 + 1 \cdot .05 + .82 \cdot .5 = .46
\]

The values assigned to (34) according to (38) are given in (39). The shaded lines will later turn out to be the source of a problem with this approach.
\[ V(W \rightarrow (S \rightarrow L))(w) = \begin{cases} 
V(S \rightarrow L)(w) & \text{if } V(W)(w) = 1 \\
E[V(S \rightarrow L)|V(W) = 1] & \text{if } V(W)(w) = 0 \\
V(L)(w) & \text{if } V(W)(w) = 1, V(S)(w) = 1 \\
E[V(L)|V(S) = 1] & \text{if } V(W)(w) = 0 \\
0 & \text{if } V(W)(w) = 1, V(S)(w) = 1, V(L)(w) = 0 \\
1 & \text{if } V(W)(w) = 1, V(S)(w) = 1, V(L)(w) = 1 \\
.82 & \text{if } V(W)(w) = 1, V(S)(w) = 0 \\
.46 & \text{if } V(W)(w) = 0 
\end{cases} \]

The distribution of these values is illustrated in Figure 3. Their expectation is again .46:

\[ P(W \rightarrow (S \rightarrow L)) = E[V(W \rightarrow (S \rightarrow L))] = \sum_{x \in [0,1]} x \cdot Pr(V(W \rightarrow (S \rightarrow L)) = x) = 0 \cdot .045 + 1 \cdot .005 + .82 \cdot .05 + .46 \cdot .9 = .46 \]

But this result is intuitively wrong. Given the scenario, the probability of the conditional should be small, rather than close to .5.

Examples like this have been used by Lance (1991) and Edgington (1991) to point out a number of similar counterintuitive predictions. In the next subsection I will discuss what the problem is and how it might be addressed.

5.1. The Values at Non-Antecedent Worlds

Intuitively, the probability of \( (34) \) should not be .46 as predicted by the Jeffrey-style value assignment. What, then, should it be? It seems like .1, the probability of \( (35d) \), would be a much better estimate. I take that probability to be that of the simple conditional in \( (41) \):

\[ \text{(41) If you strike the match and it is wet, it will light.} \]
As usual, the conditional has clearcut truth values only at those worlds at which the antecedent is true, i.e., at which the match is wet and struck. The expectation of these values, given as .1 in (35d), is uniformly distributed over the worlds at which the antecedent is false. The values are calculated as in (42); their distribution is shown in Figure 4.

\[(42) \ V((SW) \rightarrow L)(w) = \begin{cases} V(L)(w) & \text{if } V(SW)(w) = 1 \\ E[V(L) | V(SW)] = 1 & \text{if } V(SW) = 0 \\ 0 & \text{if } V(SW)(w) = 1, V(L)(w) = 0 \\ 1 & \text{if } V(SW)(w) = 1, V(L)(w) = 1 \\ .1 & \text{if } V(SW)(w) = 0 \end{cases} \]

Comparing Figures 3 and 4, it seems that the overall expectation in the latter of the values is closer to intuitions. And indeed, that expectation, given in (43), is what one would expect.

\[(43) \quad P((SW) \rightarrow L) = E[V((SW) \rightarrow L)] \\
= 0 \cdot Pr(V(SW) = 1, V(L) = 0) \\
+ 1 \cdot Pr(V(SW) = 1, V(L) = 1) \\
+.1 \cdot Pr(V(SW) = 0) \\
= 0 \cdot .045 + 1 \cdot .005 + .1 \cdot .95 \\
= .1\]
That the probability of (41) is a good estimate of that of (34) is fairly
evident. Some authors have generalized this observation and based on
it proposals for dealing with all right-nested conditionals within the
limitations of the standard probabilistic calculus. McGee (1989), for
instance, stipulates the equality in (44a). Since this definition applies
recursively, it unpacks and flattens arbitrarily deep right-embeddings
of conditionals. In the present framework, a similar effect could be
obtained by simply identifying the value assignment of one with that
of the other, as in (44b).

\[
\begin{align*}
(44a) & \quad P(\varphi \rightarrow (\psi \rightarrow \chi)) = P((\varphi \psi) \rightarrow \chi) = P(\chi | \varphi \psi) \\
(44b) & \quad V(\varphi \rightarrow (\psi \rightarrow \chi)) = V((\varphi \psi) \rightarrow \chi)
\end{align*}
\]

This “Import-Export Principle” is logically valid for the material
conditional. The diagrams in Figures 3 and 4 show why this move
would improve the predictions about the conditional in (34). The truth
values at worlds in which the match is both wet and struck are the same
in both cases. The difference lies in the worlds at which the antecedents
are false. In Figure 4, these are all the worlds at which the match is
either dry or not struck, or both.

In contrast, in Figure 3 the non-antecedent worlds are only those
at which the match is dry. Those at which it is wet but not struck
verify the antecedent, therefore the conditional there receives the value
of its consequent. Recall that the values of the consequent were given
in Figure 2 on page 31.

It would not do, of course, to stipulate somehow that the values
of the consequent should be uniformly low. For in the case of (45a),
the overall probability should arguably be high. This is again correctly
predicted if the sentence is treated as equivalent to (45b) according to
the Import-Export Principle (44b). The resulting values would then be
as shown in Figure 5 (I skip the details of the calculation).

\[
(45) \quad \begin{align*}
\text{a. } & \text{If the match is dry, then it will light if you strike it.} \\
\text{b. } & \text{If you strike the match and it is dry, it will light.}
\end{align*}
\]

To rectify this situation, it would be necessary to assign the con-
sequent (46a) different values at non-antecedent worlds depending on

\[\text{25 Notice that McGee does not also require (44a) to equal } P(\psi \rightarrow \chi | \varphi). \text{ He does not conditionalize conditional consequents on their antecedents, thus avoiding to treat them as proposition-denoting. This is what enables him to “wiggle out of the trivialization argument” (Stahlaker, 1991).}\]
whether the match is wet or not: At worlds at which it is wet, the value is that of (46b), and at worlds at which it is dry, it is that of (46c).

(46) a. If you strike it, it will light.
   b. If you strike it and it is wet, it will light.
   c. If you strike it and it is dry, it will light.

If values were assigned in this way, they would be distributed as shown in Figure 6. The values of the right-nested conditionals (34) and (45) would then come out as in Figures 4 and 5, respectively. This move looks promising, but it is at odds with what Jeffrey (1991) himself assumed about the values at non-antecedent worlds:

I take it that if $A \rightarrow C$ is to be an *indicative* conditional it must have the same value at all worlds $w$ where $A$ is false...

Jeffrey does not elaborate this claim, and the example shows that it is to blame for some problems. The shaded lines in (39) above illustrate this vividly: The value of (45a) (i.e., the consequent of (34)) at worlds at which the match is not struck is independent of whether the match is wet or not.

But it is not clear yet how an amended assignment according to (46b,c) could come about in a principled way, nor why we should not go instead with the simpler Import-Export Principle. I will discuss two arguments pertaining to these questions, one against the Import-Export Principle and one for an assignment along the lines of (46b,c). Following that, I will propose a improved version of the assignment.
5.2. THE IMPORT-EXPORT PRINCIPLE

Adams (1975) argues that the equivalence of $\varphi \rightarrow (\psi \rightarrow \chi)$ and $(\varphi \psi) \rightarrow \chi$ should not hold as a general principle. Consider the special case of $A \rightarrow (B \rightarrow A)$ and $(AB) \rightarrow A$. The latter is a logical truth, and so, given to the equivalence, would be the former. But then $B \rightarrow A$ could be inferred from $A$ and a logical truth by Modus Ponens.

Gibbard (1981, Fn. 17), who defends the Import-Export Principle, is not convinced that this prediction is all that unreasonable. However, it does seem to go against intuitions in examples like Adams'. The prediction is that (47a) and (47b) are equivalent and therefore both are logical truths.

(47) a. If the match lights, it will light if you strike it.
   b. If you strike the match and it lights, it will light.

While it is undeniable that (47b) is necessarily true, this is far less clear for (47a). Suppose again that the probability that the match lights if struck is .82 as above, and consider a situation in which it is very likely that the match is not struck but instead tossed into the camp fire, where it will light without being struck. In such a scenario, the probability of (47a) should quite clearly be less than 1. The fact that intuitions come apart so tangibly in the case of (47a) and (47b) demonstrates that the substitution is not applicable across the board as a general logical principle.

One way in which the preceding argument might be challenged is by pointing out that (34) and (47a) differ with respect to the temporal
order of the constituents: In (34), the wetness of the match presumably precedes, or is at least simultaneous with both the striking and the lighting, thus it might be suggested that the value of the conditional at worlds at which it is wet or dry, respectively, should be the one it has at that time. This, it could be argued, preserves the effect of the Import-Export Principle in this particular case but not in the case of (47a).

However, while it may be possible to make an argument along these lines for this particular case, this would not really solve the problem: There are other examples where the effect of the Import-Export Principle should obtain, but where the temporal order of the constituents is reversed like in (47a). The conditional in (48a) is such an example.

(48)  

a. If the coin comes up heads, I will lose if I bet on tails.

b. If I bet on tails and the coin comes up heads, I will lose.

Suppose the betting precedes the coin toss and (48a) is asserted prior to the betting. The temporal relations are as in (47b): The outcome of the toss and my winning or losing are determined simultaneously, and the betting (the antecedent of the embedded conditional) precedes both. But unlike (47a), the sentence in (48a) is certainly true and indeed equivalent to (48b).

Thus the temporal order of the events in question is not responsible for the failure of the Import-Export Principle in (47a,b), and therefore an appeal to temporal relations (alone) is not likely to decide when the Principle should apply.

6. Counterfactuals

In the previous section I argued that the values assigned at worlds at which the antecedents are false are responsible for the intuitively wrong probabilities of conditionals under Jeffrey’s definition. In asking what those values should be, it is natural to consult intuitions about counterfactuals. In this section I will discuss the latter to the extent that they inform the solution to the problems of the Jeffrey conditional.

Consider (49b), the counterfactual counterpart of (49a).

(49)  

a. If you strike the match, it will light.

b. If you had struck the match, it would have lit.

Clearly in the event that the match was not struck, the truth—after the fact—of the counterfactual in (49b) depends on whether it was
wet or not. Similarly, in case it was not struck but lit because it was
tossed into the fire, the fact that it lit does not make (49b) any more
of a certainty. Similarly for (48a) above: If the coin came up heads,
I would surely have lost if I had bet on tails, regardless of the fact
that the betting was done at a time at which the outcome was not yet
determined.

Thus intuitions about counterfactuals seem to accord well with what
we found the values of the indicative should be at non-antecedent
worlds. Harper (1981, fn. 18) presumably also hinted at this connection
when he noted about embedded conditionals like (47a) that “the read-
ing of the second conditional [i.e., the embedded one—SK] is implicitly
subjunctive.”

6.1. Past Predominance

In dealing with counterfactuals in time, it is often observed that the
similarity between worlds is at least to a large extent determined by
the temporal order in which facts come about. Thomason and Gupta
(1981) discussed (and ultimately dismissed) a radical implementation
of this relationship in the principle of Past Predominance:

In determining how close \([w \text{ at time } t]\) is to \([w' \text{ at time } t]\), past close-
ess predominates over future closeness; that is, the portions of \(w\)
and \(w'\) not after \(t\) predominate over the rest of \(w\) and \(w'\). (p. 301)

The simplest implementation of this idea is to re-run history from
the perspective of an earlier time at which it was still possible for the
antecedent to turn out true, then examine, from the perspective of
that point, those continuations in which it is true. In short, at a world
in which the match is not struck, the value of (49a) before the fact
should be that of (49b) after the fact. This intuition is often appealed
to (Lewis, 1979; Tedeschi, 1981; Dahl, 1997), and it is doubtlessly on
the right track. However, it does not always yield the right predictions.
The problem is that it “undoes” too much of the history of the world
of evaluation. While some posterior facts—at times after that of the
antecedent—should not bear on similarity, others should.\(^2\)

\(^2\) In probabilistic settings, the corresponding assumption that the prior prob-
ability of (49a) must equal the posterior probability of (49b) is encountered in
epistemic past (Adams, 1975) and prior propensity (Adams, 1976; Skyrms, 1981)
accounts. This view, too, faces apparently insurmountable problems (see Barker,
1998 for recent discussion). Kaufmann (2001) argues that the right way to think of
the relationship is by distinguishing between values and their expectations, just like
in the present paper. Then pairs like (49a,b), while not generally equiprobable, are
equivalent.
section I will illustrate this will some well-known examples from the literature on counterfactuals.

6.1.1. Irrelevant posterior facts
Fine (1975) presented a counterexample to Lewis’ “overall similarity” account:

The counterfactual [50] is true or can be imagined to be so.

(50) If Nixon had pressed the button there would have been a nuclear holocaust

Now suppose that there never will be a nuclear holocaust. Then that counterfactual is, on Lewis’ analysis, very likely false. For given any world in which antecedent and consequent are both true it will be easy to imagine a closer world in which the antecedent is true and the consequent false. For we need only imagine a change that prevents the holocaust but that does not require such a great divergence from reality.

Fine’s example suggests that the appeal to overall similarity alone, understood as similarity with respect to all facts throughout history, is not quite right. Posterior facts should be discounted. Small changes may cause vast divergences between worlds, but what counts for similarity is only the history up to the small change required to make the antecedent true.27

6.1.2. Relevant posterior facts
Contrary to the preceding example, the following one, which already appeared in the right-nested conditional (48) above, shows that in some cases posterior facts do affect the interpretation of the counterfactual. Slote (1978) attributes the example to Sidney Morgenbesser. I cite it here in the version of Bennett (1984):

At \( t_1 \) I bet that when the coin is tossed at \( t_2 \) it will come up heads; and in the upshot it does just that; but this is a purely chance event, with no causally sufficient prior conditions. Now consider the conditional

(51) If I had bet on tails at \( t_1 \) I would have lost.

---

27 Lewis (1979) responded by appealing to an intricate hierarchy of “miracles”: The “small” miracle which would cause Nixon to press the button is outweighed by the “large” miracle that would be required to restore similarity in particular fact after the button is pushed, since all traces and consequences of Nixon’s hypothetical act would have to be obliterated. Many authors have complained that Lewis’ notion of “miracle” is too vague to test the predictions of the theory (cf. Nute 1980, 1984).
I agree with Bennett’s judgment and that of his informants that (51) should come out true.\textsuperscript{28} However, this is not predicted by Past Predominance as it stands. The problem is the same as before: In worlds which grow out of the same history as the actual one at $t_1$, the chance process of tossing the coin has not yet run its course. The prior chance of tails is not zero, not even small, hence the counterfactual is wrongly predicted to be false.

Notice also that these intuitions cannot be explained by the probabilities of the scenario: For instance, imagine that instead of one fair coin there were two, used depending on my bet. Then a different coin would have been used if I had bet on tails. Since it, too, would have been fair, the probabilities are the same; but the counterfactual in (51) is now true. It is stochastic dependencies as causal ones that are behind these intuitions.\textsuperscript{29}

6.1.3. \textit{Tichý’s Puzzle}

The following little story, due to Tichý (1976) and frequently discussed in this connection (Veltman, 1985, and others), is a useful example which, when viewed from various different angles, summarizes the points made so far.

(T1) Consider a man—call him Jones—who is possessed of the following disposition as regards wearing his hat. Each morning he flips a coin before he opens the curtains to see what the weather is like. ‘Heads’ means he is going to wear his hat if the weather is fine, ‘tails’ means he is not going to wear his hat in that case. Bad weather, however, invariably induces him to wear a hat.

This morning, ‘heads’ came up when he flipped the coin; furthermore, the weather is bad, so Jones is wearing his hat; and

\textsuperscript{28} One reviewer takes issue with this intuition of Bennett’s, his informants, and mine. According to the reviewer, the following sentence (a) is true in the situation described, simply due to the fact that the outcome was “genuinely chancy,” i.e., that both sentences in (b) were true at $t_1$.

a. If I had bet on tails I might have won.

b. It might land heads. / It might land tails.

c. I might have bet on tails and won.

I think it is an open empirical question how easy the reviewer’s reading for (a) is to come by. I think (c) is true in the situation by virtue of the fact that the outcome was chancy, but things are much less clear for (a). The reviewer’s suggestion deserves further investigation; meanwhile, however, I suspect that most speakers would agree that (a) strongly suggests that the outcome somehow causally depended on the bet.

\textsuperscript{29} Adams (1975, p. 129) already pointed out the importance of distinguishing stochastic from causal dependence in cases like this.
(52) If the weather were fine, Jones would be wearing his hat.

Had the weather been fine, the coin would nevertheless have come up heads. Thus (52) is correctly predicted true under past predominance, since Jones tossed the coin before he knew what the weather was like. Consider now the following minor variant of the story:

(T2) This morning, Jones was too impatient to toss the coin first. He opened the curtains and saw that the weather was bad. But he tossed the coin all the same, since doing so is a time-honored habit of his. The coin came up ‘heads’. Later in the day, he is wearing his hat.

Clearly in this case, (52) should still be true. However, this is no longer predicted by Past Predominance: When Jones found out about the weather, he had not yet tossed the coin. A re-run of history from that point onward would involve a different toss of the coin, one whose outcome could well turn out different. Thus it would be predicted, contrary to intuitions, that if the weather were fine, Jones might not be wearing his hat.

This prediction would be correct, however, if John’s decision whether or not to toss the coin depended on the weather, as in the following variant:

(T3) Jones always opens the curtains first. If the weather is bad, he will wear his hat. If it is good, he tosses the coin. This morning the weather was bad, so he is wearing his hat.

If the weather had been fine, he would have tossed the coin to decide whether to wear his hat, so he might not wear it.

What if he tosses the coin no matter what, but the consequences of the toss (i.e., its effect on John’s behavior) depend on the weather? Consider the following variation on Tichý’s theme:

(T4) Jones always opens the curtains and then tosses the coin. In bad weather, the coin decides which of his two bad-weather hats he will wear—‘heads’ for blue, ‘tails’ for black. In good weather, the toss decides whether he will wear his straw hat (‘heads’) or none at all (‘tails’). Today the weather is bad and the coin came up ‘heads’, so he is wearing his blue hat.

Clearly (52) is once again true in this scenario. If the weather were fine, the outcome would still have been ‘heads’ and it would, jointly with the weather, have induced Jones to wear his straw hat.
Notice that in none of these cases did we assume that the probabilities of heads and tails, if a toss is made, depend in any way on the weather. Notice also that in (T1), (T3) and (T4) what makes the counterfactual true is not the fact that a toss was made, but rather the particular "toss token," along with its outcome. For suppose we are told in addition that in bad weather Jones tosses the coin in the living room, whereas in good weather he tosses it on the balcony. Now it is no longer the case that in good weather it would have been the very same toss which would have decided, so the possibility that it might have come up 'tails' is again open, and (52) is intuitively false.

6.2. CAUSAL INDEPENDENCE

Examples like those discussed above suggest that if the unraveling and re-running of history is to lead to the right predictions about the values of counterfactuals, simple temporal precedence must be augmented with further restrictions on which courses of events are to be considered as alternatives to the actual one. Some facts that are settled after the reference time of the antecedent must be held constant.

What are the facts that should be held constant? Obviously those that would not be different if the truth value of the antecedent had been otherwise. But this is itself a counterfactual. This circularity has plagued theories of counterfactuals ever since it was first made explicit by Goodman (1947). The notion of relative similarity employed by Stalnaker and Lewis is way of breaking it, or rather sidestepping it, by using "a model theory with room for wired-in answers to various counterfactual questions" (Jeffrey, 1991, p. 162).

Another way of avoiding the circularity takes causal relations as basic: Some processes are not causally affected by the antecedent, and their outcomes are to be carried over in choosing alternative worlds. The relation of being (or not being) "causally affected," simply an assumption speakers seem to be making, remains unanalyzed and is given as part of the model. Ultimately, therefore, the avoidance of circularity is only apparent, since no analysis of causality seems to be forthcoming that would not in turn refer to counterfactuals. But taking causality as basic seems appropriate if the goal is a semantic analysis of conditionals, rather than a metaphysical analysis of causality.

---

30 See Mårtensson (1999) and Bennett (2003) for a related and more detailed analysis in terms of causal "trails" and "ramps," "early departure" and other notions.
6.3. What causal relations relate

In order to integrate causal relations into the model, a few more precisifications are in order. First, the causal relations I have in mind hold between event tokens rather than types.\(^{31}\) The former is typically described by episodic sentences like (53a), the latter by generics like (53b).

(53) a. Her lack of exercise caused her heart attack.
    
    b. Lack of exercises causes heart attack.

Although pairs such as (53a,b) superficially appear to make the same claim, the relations referred to have different properties. It is natural to assume that causal relations between event tokens are transitive, asymmetric, and irreflexive. The same does not hold for event types (Davis, 1988).

Related to this distinction is the second question, how to encode the relation formally. In most probabilistic theories, the formal analog of the claim that ‘C causes E’ is the statement that ‘C increases the likelihood of E (under certain circumstances)’, accompanied by the condition that the high probability E does not come about in other ways or cannot be explained otherwise. Suppes (1970) applied a definition of this kind indiscriminately to both token and type causation, but subsequent authors came to agree that it is more appropriate for the latter than the former (see Davis, 1988 for an overview).

Instead, to formalize causal relations between event tokens, I take causation to be a deterministic relation between events and probabilities: The statement that ‘C causes E’ asserts that the probability of E depends on whether or not C occurs. To adapt Hausman’s (1998) example, “rather than regarding the match’s lighting (if it does) as probabilistically caused, one should regard the probability that it lights as deterministically caused” (p. 201). Hausman lines up an array of further good arguments for this view. I adopt it here without further discussion.

\(^{31}\) This talk of “events” as the carriers of the relation is somewhat loose and sidesteps a number of complications. Hausman (1998) settles on “tropes,” parts or aspects of events, of which he writes: “A trope is a located value of a variable or an instantiation of a property at a place and time” (p. 26). I will continue to speak of events for now, and switch to indicator variables in the formal implementation.
6.4. Implementation

Events, the bearers of causal relations, are not directly represented in a model defined as above, in which worlds have no structure other than a temporal succession of states. However, for each event there is a characteristic function which identifies the set of those worlds in which it occurs. The language may contain a sentence that is true if and only if the event occurs, in which case the denotation of the sentence coincides with this characteristic function. There is no a priori reason to restrict the possible causes and effects to those that can be named, however.

An effect is not an event, but the probability of an event. The event itself, like its cause, is represented by the characteristic function of the proposition that it occurs. The expectation of this function is the chance that the event has of occurring. It can be read off the model: Each world at each time \( t \) is a member of a set \( [w]^2 \) of historical alternatives. The chance of the event is \( Pr(\text{the event occurs}|[w]^Z) \).

The relation of causal dependence is defined solely in terms of the variables involved, but separate from the probability distribution (see also Woodward, 2001): A statement of the form ‘\( X \) causally affects \( Y \)’, where \( X \) and \( Y \) are random variables, asserts that the value of \( X \) causally affects the expectation of \( Y \).

In Definition 13, a subset of all random variables is singled out and collected in a set \( \Phi \) of “causally relevant” variables in the model.

**Definition 13 (Causal chance model)**

A causal chance model is a structure \( \langle W, T, \prec, \approx, Pr, \Phi, \prec \rangle \), where \( \langle W, T, \prec, \approx, Pr \rangle \) is a chance model (cf. Definition 11, page 27) and \( \langle \Phi, \prec \rangle \) is a set of random variables on \( W \), where \( \prec \) is a strict partial order.\(^{32}\)

The idea is that \( \varphi \prec \varphi' \) means ‘\( \varphi \) causally affects \( \varphi' \)’ in the above sense. In the interpretation of conditionals, in fact, it will be more useful to identify those variables which are ‘causally unaffected’ by a given variable. These are all its non-descendants in the causal order.

**Definition 14 (Causal independence)**

Given a causal structure \( \langle \Phi, \prec \rangle \), for all \( \varphi, \varphi' \in \Phi \): \( \varphi' \) is causally independent of \( \varphi \) if and only if \( \varphi \not\prec \varphi' \).

As in the case of probabilities, this causal structure can be given a subjective or an objective interpretation. It may be reasonable to

\(^{32}\) Transitivity is commonly assumed for token causation (Davis, 1988, p. 146), but it only holds for the relation itself, not necessarily for the probabilistic “impulse” transmitted along a causal chain; the latter may be canceled out by the values of intermediary variables; cf. also Hausman (1998, p. 204); Pearl (2000, p. 237).
assume that objective chance comes with an “all-embracing Bayesian net” (Spohn, 2001), a frame “containing all variables needed for a complete description of empirical reality” (p. 167). If, on the other hand, the evaluation in the model is to serve as a realistic reflection of the interpretations carried out by actual speakers, the set of variables that are considered relevant in any given instance would probably be rather sparse.

7. Value assignments

With an explicit encoding of causal dependencies as in Definition 13, we can now take their effect on the interpretation of conditionals into account by making the value assignment sensitive to them. Given an ordered set $\langle \Phi, \prec \rangle$ of relevant variables as defined above, the definition in (27) on page 22 is changed to (54):

$\begin{align*}
V(\varphi \rightarrow \psi)(t)(w) &= \begin{cases} 
V(\psi)(t)(w) & \text{if } V(\varphi)(t)(w) = 1 \\
E[V(\psi)(t)|V(\varphi)(t) = 1, [w]_I^\infty, X = X(w)] & \text{for all } X \in \Phi \text{ s.t. } V(\varphi)(t) \not\prec X \\
0 & \text{if } V(\varphi)(t)(w) = 0
\end{cases}
\end{align*}$

Recall from the discussion in Section 5 that in order for the predictions about embedded conditionals to come out right, the value of (55a) should be that of (55b) at worlds at which the match is wet and that of (55c) at worlds at which it is dry:

(55)

a. If you strike the match, it will light.

b. If you strike the match and it is wet, it will light.

c. If you strike the match and it is dry, it will light.

According to (54), these are indeed the values of the conditional, provided that $\langle \Phi, \prec \rangle$ holds the information that the wetness of the match does not causally depend on the striking.\footnote{Notice again that this is orthogonal to any stochastic dependence between two: Perhaps you won’t strike the match unless it is dry, so your not striking it is evidence that it is wet. But that has no bearing on the question of what would have happened if you had struck it in the event that it is wet.} Let us see how in this case the distribution of the values of (55a) comes out as shown in Figure 6 on page 37, which I argued is correct.

\begin{itemize}
  \item [55] a. If you strike the match, it will light.
  \item [55] b. If you strike the match and it is wet, it will light.
  \item [55] c. If you strike the match and it is dry, it will light.
\end{itemize}
Figure 7. Causal dependencies in (55)

The structural dependencies among the relevant variables are simple: Whether the match lights depends both on its wetness and on whether it is struck or not; however, the wetness and the striking are causally independent from each other. These facts may be represented schematically as in Figure 7. By the above definition, it follows that four different values are assigned to the conditional in (55a), since the value at worlds at which the match is not struck depends on whether it is wet or not. The assignment is as shown in (56). See Figure 6 on page 37 for illustration.

\[
V(S \rightarrow L)(w) = \begin{cases} 
V(L)(w) \text{ if } V(S)(w) = 1 \\
E[V(L)|V(S) = 1, V(W) = 0] \\
E[V(L)|V(S) = 1, V(W) = 1] \\
0 \text{ if } V(S)(w) = 1, V(L)(w) = 0 \\
1 \text{ if } V(S)(w) = 1, V(L)(w) = 1 \\
.9 \text{ if } V(S)(w) = 0, V(W)(w) = 0 \\
.1 \text{ if } V(S)(w) = 0, V(W)(w) = 1 
\end{cases}
\]

The expectation of these values is the probability of (55a), predicted to be high:

\[
P(S \rightarrow L) = 0 \cdot Pr(V(S) = 1, V(L) = 0) + 1 \cdot Pr(V(S) = 1, V(L) = 1) + .9 \cdot Pr(V(S) = 0, V(W) = 0)
\]
+.1 \cdot Pr(V(S) = 0, V(W) = 1)
= 0 \cdot .09 + 1 \cdot .41 + .9 \cdot .45 + .1 \cdot .05
= .82

Turning to the right-nested conditional (58a), its values are now different because the values of its consequent are distributed differently.

(58) a. If the match is wet, then if you strike it it will light.
    b. If the match is dry, then if you strike it it will light.

Notice in particular the gray area in (59), which highlights the contrast with the problematic earlier formula in (39) on page 33. As a result of this change, the values at those worlds at which the match is not wet, calculated as the conditional expectation of the values of the consequent over those worlds at which it is wet, is low as well.

(59) $V(W \rightarrow (S \rightarrow L))(w)$

\[
= \begin{cases} 
V(S \rightarrow L)(w) & \text{if } V(W)(w) = 1 \\
E[V(S \rightarrow L)|V(W) = 1] & \text{if } V(W)(w) = 0 \\
V(L)(w) & \text{if } V(W)(w) = 1, V(S)(w) = 1 \\
E[V(L)|V(S) = 1, V(W) = 1] & \text{if } V(W)(w) = 0 \\
E[V(S \rightarrow L)|V(W) = 1] & \text{if } V(W)(w) = 0 \\
0 & \text{if } V(W)(w) = 1, V(S)(w) = 1, V(L)(w) = 0 \\
1 & \text{if } V(W)(w) = 1, V(S)(w) = 1, V(L)(w) = 1 \\
.1 & \text{if } V(W)(w) = 1, V(S)(w) = 0 \\
.1 & \text{if } V(W)(w) = 0
\end{cases}
\]

A diagram of the resulting values already appeared in Figure 4 on page 34. It is obvious from that picture that the expectation of the values will be small. This is indeed the case: The probability of (58a) under the new value assignment is much more in line with intuitions, as (60) shows.

(60) $P(W \rightarrow (S \rightarrow L))$

\[
= 0 \cdot Pr(V(W) = 1, V(S) = 1, V(L) = 0) \\
+ 1 \cdot Pr(V(W) = 1, V(S) = 1, V(L) = 1) \\
+.1 \cdot Pr(V(W) = 1, V(S) = 0)
\]
\[ +.1 \cdot Pr(V(W) = 0) \]
\[ = 0 \cdot .045 + 1 \cdot .005 + .1 \cdot .05 + .1 \cdot .9 \]
\[ = .1 \]

On the other hand, the values and their expectations are high for the conditional in (58b). The value assignment is as follows:

\[(61) \quad V(\overline{W} \rightarrow (S \rightarrow L))(w) \]
\[ = \begin{cases} 
V(S \rightarrow L)(w) & \text{if } V(W)(w) = 0 \\
E[V(S \rightarrow L)|V(W) = 0] & \text{if } V(W)(w) = 1 \\
V(L)(w) & \text{if } V(W)(w) = 0, V(S)(w) = 1 \\
E[V(L)|V(S) = 1, V(W) = 0] & \text{if } V(W)(w) = 1 \\
0 & \text{if } V(W)(w) = 0, V(S)(w) = 1, V(L)(w) = 0 \\
1 & \text{if } V(W)(w) = 0, V(S)(w) = 1, V(L)(w) = 1 \\
.9 & \text{if } V(W)(w) = 0, V(S)(w) = 0 \\
.9 & \text{if } V(W)(w) = 1 
\end{cases} \]

Again, the shaded area marks the effect of the new definition, according to which the values of the conditional \( S \rightarrow L \) depend on whether the match is wet or dry at the world of evaluation. The distribution of these values is as shown in Figure 5 on page 36.

The probability of (58b), the expectation of the values defined in (61), is accordingly high, as desired:

\[(62) \quad P(\overline{W} \rightarrow (S \rightarrow L)) \]
\[ = 0 \cdot Pr(V(W) = 0, V(S) = 1, V(L) = 0) \]
\[ + 1 \cdot Pr(V(W) = 0, V(S) = 1, V(L) = 1) \]
\[ + .9 \cdot Pr(V(W) = 0, V(S) = 0) \]
\[ + .9 \cdot Pr(V(W) = 1) \]
\[ = 0 \cdot .045 + 1 \cdot .405 + .9 \cdot .45 + .9 \cdot .1 \]
\[ = .9 \]
8. Further issues

In this paper I dealt with simple and right-nested predictive conditionals and argued that a causal account, which is independently needed to handle for counterfactuals, corrects the problematic predictions of the random-variable approach. I believe that this improved account brings us closer toward a unified probabilistic semantics for the various classes of conditionals (and ultimately not just those), but some issues that I did not touch on in this paper need to be addressed before such a theory will take shape. I will conclude by summarizing and re-stating the main issues that are left for future work.

As Stalnaker and Jeffrey (1994) showed, the random variable approach can be extended to a language containing conjunctions and arbitrary embeddings of conditionals, including ones with conditional antecedents. However, this general case requires a more complex model theory. Stalnaker and Jeffrey borrow the necessary apparatus from van Fraassen’s (1976). Unfortunately, the predictions about the probabilities of the more complex sentences thus covered are again counterintuitive in some cases (Lance, 1991; Edgington, 1991). These problems might be overcome by modifying the account along similar lines as the proposal in this paper, but considerable technicalities are involved in spelling this out.

The restriction to predictive conditionals and their counterfactual counterparts was, as I mentioned above, for the sake of simplicity. Both the problems discussed here and their solution in fact apply as well in the case of epistemic conditionals and subjective probability. But while the solution works for each of these interpretations, a theory in which both are combined requires extra care in setting up the relationship between them.

Finally, I already made clear that the value assignment I proposed in Section 7 relies on an assumption that clearly cannot be sustained in general: that some set of antecedent worlds with positive probability can be accessed from every non-antecedent world while keeping constant the values of all variables that do not causally depend on the antecedent. The assignment in (54) does not define a value if the values of those variables at the world of evaluation are such that the antecedent has zero probability. More generally, this paper only introduced restrictions on the accessibility of antecedent worlds. That restrictions are needed was shown by the examples; however, assuming that patches of undefined values in the set of historical alternatives are undesirable, the restrictions on accessibility must be counterbal-
anced by the desire to ensure that the values of conditionals be defined everywhere if they are defined anywhere.\textsuperscript{34}

I leave each of these issues for another occasion. Trying to discuss them all would have made this paper even more lengthy and unwieldy.

Whether the appeal to causal relationships truly solves the problem or merely begs the question may be debatable. However, it is interesting to know that once this move is made, the relationship between counterfactual and indicative conditionals becomes very straightforward and predictions about embedded conditionals become sensible. This should be seen in the context of the rising recognition of the role of causality in other areas, notably Artificial Intelligence.\textsuperscript{35}

\section*{References}


\textsuperscript{34} This problem was addressed by Skyrms (1984, 1988, 1994), cited above, in his set selection account. There, too, the values of conditionals are calculated locally within the cells of partitions induced by causal background factors. Cells in which the antecedent does not have positive probability are merged with larger ones in which it does. How exactly this merger proceeds is not very clear; Skyrms does not assume that it is guided by an order on the background factors, as would be natural in the present proposal. Here, the corresponding solution would consist in conditioning on fewer variables in (54).

In Section 2.4 I mentioned that Skyrms' account is in some respects similar to mine. This similarity breaks down in other areas; for instance, for counterfactuals he assumes a prior propensity account, which is vulnerable to Barker’s (1998) objections.

\textsuperscript{35} Readers familiar with probabilistic methods in knowledge representation will recognize a connection between the approach advocated here and the theory of causal Bayesian networks (Cowell et al., 1999; Spirtes et al., 2000; Pearl, 2000, and elsewhere), on which I did not elaborate in this paper.


kaufmann@northwestern.edu
Draft - do not cite. Comments welcome.
July 6, 2003
Spohn, W.: 2001, ‘Bayesian Nets are All there is to Causal Dependence’. In (Galavotti et al., 2001), pp. 157–172.
Thomason, R. H.: 1984, ‘Combinations of Tense and Modality’. In (Gabbay and Guenthner, 1984), pp. 135–165.
Figure 1
Figure 3
Figure 4
Figure 5
Figure 6