Probabilities of Conditionals

Stefan Kaufmann
Stanford University

1. Introduction

It is a well-known fact that the material conditional familiar from classical logic is not a good translation of English constructions involving *'if ... then ... ,'* even though it is the best *truth-functional* connective to serve that purpose. Work on the semantics of modality and counterfactual conditionals (Lewis, 1973; Kratzer, 1979) inspired an alternative treatment of indicative conditionals in terms of restricted quantification over possible worlds (Kratzer, 1986, 1991). Given a *conversational background*—a contextually given set of possible worlds—the *antecedent* (that is, the denotation of the protasis) of a sentence like

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

restricts that set of worlds to those in which strike the match, and the *consequent* (the denotation of the apodosis) is interpreted as a universally quantified statement asserting that the match lights in *all* of those worlds. More precisely, symbolizing the background as "K," we can say:

Definition 1 *A conditional 'If A then C' is true with respect to a set of worlds K if and only if for all worlds in K in which A is true, C is true.*

This paper addresses two questions left open by this approach:

(2)   a.   What is the force of the quantification?
        b.   What is the source of the "background" set of worlds?

The first is motivated by the fact that if, as in Definition 1, the quantification over the relevant set of worlds is *universal*, then a number problematic inference patterns remain unaccounted for. According to Definition 1, a conditional 'If A then C' is true with respect to a background set K if and only if the corresponding material conditional $A \supset C$ is true at all worlds in K. Thus if, for instance, K is taken to represent the beliefs of the speaker, then belief in the natural-language conditional is predicted to be belief in the material conditional. In Section 2 I use a number of examples to cast doubt on that, and in Section 3 I argue the merits of a *probabilistic* alternative.

The second question concerns the origins of the set of worlds providing the "background" (K in Definition 1.) It is often identified with the set of those worlds that are compatible with what is *known* by the speaker, all participants in the conversation, or some hypothetical agent. Kratzer (1991) considers this epistemic

interpretation the most basic and common one, although she also mentions different kinds of modality (deontic etc.) for which different sets may be relevant.

Different epistemic states may give different values to the same conditional. This raises the question as to whether conditionals have *objective* truth values independently of anyone's beliefs, defined solely in terms of the *facts* of the world. If they do, question (2b) must be addressed in addition to (2a). If they do not, then conditionals can only be treated as statements about *beliefs* and it is not clear whether the belief in a conditional is the belief that it is *true*.

Section 4 relates both the origins of the set of worlds and the probabilistic alternative to universal quantification to a theory of objective chance. Finally, Section 5 suggests further refinements of the account, drawing on intuitive connections between conditional predictions and counterfactual conditionals.

This paper is mainly concerned with the values of *predictive conditionals*, illustrated in (3a). *Epistemic conditionals* like (3b) are beyond its scope, and *counterfactuals* such as (3c) are only briefly touched upon. For a more comprehensive account, see Kaufmann (2001).

(3)   a. If you strike the match, it will light.   [*predictive*]

           b. If you struck the match, it lit.   [*epistemic*]

           c. If you had struck the match, it would have lit.   [*counterfactual*]

2. Data: Inference patterns

In this section I review some inference patterns which demonstrate that the material conditional is not an appropriate connective to model our everyday use of conditionals.[1] As shown in Section 1, this inadequacy of the material conditional carries over to an account in terms of universal quantification over worlds.

Each of the following four subsections will start with a valid inference pattern involving the material conditional, then present a linguistic example to show that the same pattern is not valid with the "natural" conditional.

2.1. *Strengthening the antecedent*

(4) $$\frac{A \supset C}{AB \supset C}$$

However: Far from contradicting (5a), (5b) is a perfectly natural continuation. Thus the two sentences in (5c) are consistent.

(5)   a. "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.

          b. But if the burglars learn how to trick this alarm system, that is a rise in price, because the quality advantage will be eroded. [nyt961217.0474]

          c. $A \rightarrow C$ and $AB \rightarrow \overline{C}$.

2.2. *Contraposition*

$$(6) \qquad \frac{A \supset C}{\overline{C} \supset \overline{A}}$$

However: (7a) may be true while (7b) is false. One may assent to the former while rejecting the latter.

(7)   a.   If you're a high achiever, it takes a long time to get recognition."

[nyt961130.0128]

b.   ??If it takes a short time to get recognition, you're a low achiever.

c. $A \rightarrow C$. But *not* $\overline{C} \rightarrow \overline{A}$.

2.3. *Vacuous truth*

$$(8) \qquad \frac{\overline{A}}{A \supset C}$$

However: The truth of (9a) does not ensure the truth of (9b).

(9)   a.   The flood crest won't reach the levels projected.

b.   ??If the flood crest reaches the level projected, much of the city will be under water.

c. $\overline{A}$. But *not* $A \rightarrow C$.

2.4. *Hypothetical Syllogism*

$$(10) \qquad \begin{array}{c} B \supset C \\ A \supset B \\ \hline A \supset C \end{array}$$

However: It is possible to assent to both of (11a,b), yet reject (11c).

(11)   a.   If I quit my job, I can't afford my apartment.

b.   If I win a Million, I'll quit my job.

c.   ??If I win a Million, I can't afford my apartment.

d. $B \rightarrow C$ and $A \rightarrow B$. But *not* $A \rightarrow C$

It has been pointed out (Adams, 1975; Kratzer, 1986, and elsewhere) that counterexamples to Transitivity typically rely on changes in the conversational background and premises which would not in fact be tenable simultaneously with respect to the *same* background. In (11), too, this can be exposed by changing the order of the premises: (12a,b) together are much less plausible than (11a,b), a fact which is attributable to the tendency to interpret (12b) as the *modally subordinated* (12c).

(12)   a.   If I win a Million, I'll quit my job.

        b.   ??If I quit my job, I can't afford my apartment.

        c.   ??If I *win a Million and* quit my job, I can't afford my apartment.

(13) $A \rightarrow B$

        $\dfrac{AB \rightarrow C}{A \rightarrow C}$

This is in line with the more general fact that the inference illustrated in (13) does indeed seem valid. This, too, needs to be explained.

3. Probability

The account I am going to propose is a probabilistic one. This section introduces some relevant notions and motivations, as well as a first step towards a formal implementation.

3.1. *What is probability?*

Like truth values, probabilities are numbers attached to propositions and, by extension, to sentences. Unlike truth values, probabilities can be *intermediate*, between 0 and 1 (inclusive.) What it means for such a value to be short of 1 depends on what it would mean for it to be 1: If "1" means "true," then a value short of 1 signifies an objective *chance*. If "1" means "known," a value short of 1 signifies a degree of (subjective) *credence* or epistemic support. Both of these interpretations are related—subjective credence can be thought of as an estimate of objective chance—and both follow the same probabilistic calculus. Here I focus on the *objective* values of conditionals.

The probabilistic calculus defines the values of propositions and compounds thereof. In addition, it defines, for each pair of propositions X and Y, the *conditional probability* of Y, given X, whenever the probability of X is positive. This can be paraphrased as "the probability Y will (would) have if X is (were) the case."

3.2. *Why probability?*

Many authors have been intrigued by the parallelism between the informal characterization of conditional probability and the natural-language conditional. Compare (14a,b):

(14)   The value of *'If A then C'* is that of ...

        a.   *'Either A is false or C is true, or both.'*

        b.   *'C is true, assuming / given / provided that A is true.'*

To many, (14b) just "sounds right." Van Fraassen (1976) made one of the many statements to this effect:

> [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?   (pp. 272–3)

Besides this "gut feeling," however, there are powerful theoretical arguments in favor of the probabilistic account. Adams (1965, 1975, 1998) explored its consequences in a theory of probabilistic inference, central to which are the following two assumptions:

(15)   a.   The probability of a conditional *'If A then C'* is the conditional probability of C, given A; and

      b.   Inference preserves (high) probability: An inference pattern is probabilistically valid if and only if it is impossible for the premises to be highly likely while the conclusion is not highly likely.[2]

The resulting system of inference has several desirable properties; the two most relevant in the present context are the following:

(16)   a.   All probabilistically valid inferences are classically valid.

      b.   Some classically valid inferences are *not* probabilistically valid: Strengthening the antecedent, Contraposition, Vacuous truth, Transitivity.

In addition, while Transitivity is invalid, the variant involving modal subordination, illustrated in (12), *is* valid, as desired.

Aside from other advantages, these properties—its conservativity with respect to the classically valid inferences and its predictions as to which inferences involving conditionals are invalid—strongly recommend a probabilistic semantics as a superior model of the way we actually use and reason with conditionals. But can it be integrated with the familiar possible-worlds semantics?

3.3. *Probabilities of propositions*

The primary use of probabilities is in modeling uncertainty. Uncertainty in turn is modeled by spelling out all the possibilities—possible worlds in the familiar sense. Let W be such a set of worlds. A *probability distribution* is a function Pr from subsets of W to real numbers, satisfying the conditions in Figure 1.[3]

For all subsets X of W, $Pr(X)$ is the probability that the actual world is in X. Thus $Pr(W)$ is merely the probability that we are in a world. The third condition ensures, roughly speaking, that "the larger the set, the larger the probability." The conditional probability $Pr(Y|X)$ of Y, given X, is the amount of Y-worlds among the X-worlds, as shown in Figure 2.
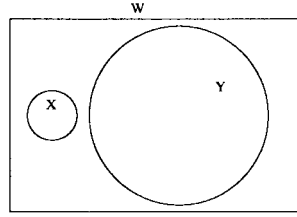
$$Pr(W) = 1$$
$$0 \leq Pr(X) \leq 1 \text{ for } X \subseteq W$$
$$Pr(X \cup Y) = Pr(X) + Pr(Y)$$
$$\text{if } X \text{ and } Y \text{ are disjoint}$$

Figure 1: Probability
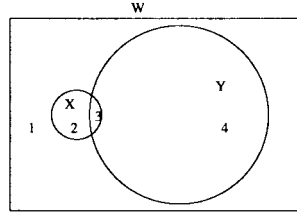
$$Pr(Y|X) = \frac{Pr(3)}{Pr(2) + Pr(3)}$$
$$\text{but undefined if X is empty}$$

Figure 2: Conditional probability

Thus the definition of conditional probabilities captures the same intuition that is behind the treatment of conditionals in terms of restricted quantification. The difference is that conditional probability admits of *degrees*: Whereas anything short of "1" is "false" on the logical account, a high conditional probability may still render a sentence "true enough."

3.4. *Probabilities of sentences*

The function Pr is defined over sets of worlds. At issue here, however, are the values of *sentences*. Atomic sentences receive their truth values at individual worlds with reference to the *facts* at those worlds. A valuation function V is defined for all pairs of atomic sentences and worlds, and extended to all truth-functional compounds of sentences not containing conditionals, as follows:

(17) $V(A)(w) \in \{0, 1\}$

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

(19) $V(\varphi \wedge \psi)(w) = V(\varphi)(w) \cdot V(\psi)(w)$

Based on V and Pr, we now define a measure P on sentences: $P(A)$ is the *expectation*, or weighted average, of the values of $V(A)$, where the weights are the probabilities, for each value x in the range of $V(A)$, that $V(A)$ assigns x. I will abbreviate expressions of the form "$\{w \in W | V(\varphi)(w) = x\}$" as "$\{V(\varphi) = x\}$." Then[4]

$$(20) \qquad P(A) =_{df} E[V(A)] = \sum_{x \in range(V(A))} x \cdot Pr(\{V(A) = x\})$$

$$= 1 \cdot Pr(\{V(A) = 1\}) + 0 \cdot Pr(\{V(A) = 0\})$$

$$= Pr(\{V(A) = 1\})$$

Thus $P(A)$, the value of A under Pr and V, is just the probability that A is true. As in the case of V, this extends straightforwardly to truth-functional compounds of atomic sentences. We can now also define *conditional* values of sentences in terms of *conditional expectations*:

$$(21) \qquad P(C|A) =_{df} E[V(C)|\{V(A) = 1\}]$$

$$= \sum_{x \in range(V(C))} x \cdot Pr(\{V(C) = x\}|\{V(A) = 1\})$$

$$= Pr(\{V(C) = 1\}|\{V(A) = 1\})$$

Given the overall goal of relating the probabilities of conditionals to their truth values on the one hand and to conditional probabilities on the other, the task now is to define $V(A \rightarrow C)$ in such a way that $P(A \rightarrow C)$ equals $P(C|A)$ regardless of Pr.

3.5. *Probabilities of conditionals*

So far the values of conditionals at individual worlds are not defined. What should $V(A \rightarrow C)(w)$ be?

This question has been at the center of considerable debate in the philosophical literature. What would appear to be the most straightforward definition is not available: It is not generally possible to distinguish, once and for all, the worlds at which the conditional is *true* from those at which it is *false*. The following is merely an indication of one aspect of the problem; there is no room here to discuss it in full detail.

There are no "conditional propositions," in the following sense: For two propositions X and Y, there is no proposition Z (except for certain pathological cases) such that the equality $Pr(Z) = Pr(Y|X)$ is guaranteed for all Pr under which both are defined. Intuitively, this is because $Pr(Y|X)$ depends only on $Pr(X)$ and $Pr(XY)$ while the Z we seek to identify must be allowed to include worlds outside X. To ensure the equivalence of $Pr(Z)$ and $Pr(Y|X)$, there must be worlds outside X that are in Z when $Pr(Y|X)$ is large and not in Z when $Pr(Y|X)$ is small. But then membership in Z cannot be determined once and for all, hence Z is not a proposition. Consequently, a definition identifying $\{V(A \rightarrow C) = 1\}$ with Z is not forthcoming.

The *triviality results* of Lewis (1976, 1986b) show that one apparently plausible way of addressing the issue—by stipulating that the truth or falsehood of the

consequent is sufficient for the truth or falsehood, respectively, of the conditional—is not viable. Despite the importance of these results in the history of the field, I will not say more about them here. The approach presented below is not vulnerable to them.[5]

4. The values of conditionals

The account I present here has precursors in the work of van Fraassen (1976) and Jeffrey (1991). The common idea is to assign *intermediate values*, between 0 and 1, to conditionals at worlds.[6] The advantages of my treatment that I want to deal with in some detail are (i) its metaphysical foundation, relating the values to objective chances and thus to the *facts* at the world of evaluation; and (ii) the values predicted for complex conditionals, discussed in Section 5.

4.1. *Intermediate values*

The definition distinguishes between those worlds at which the antecedent is true and those at which it is false. At the former, the values coincide with those of the material conditional:

(22) $$V(A \rightarrow C)(w) = V(C)(w) \text{ if } V(A)(w) = 1$$

This defines values for all worlds at which the antecedent is true. (23) is the *conditional expectation* of $V(A \rightarrow C)$ over the set of those worlds:

(23) $$E[V(A \rightarrow C)|\{V(A) = 1\}] = E[V(C)|\{V(A) = 1\}]$$
$$= Pr(\{V(C) = 1\}|\{V(A) = 1\})$$

The idea is to let this conditional expectation be the value assigned by $V(A \rightarrow C)$ at those worlds where the antecedent is false:

(24) $$V(A \rightarrow C)(w) = \begin{cases} V(C)(w) \text{ if } V(A)(w) = 1 \\ E[V(A \rightarrow C)|\{V(A) = 1\}] \text{ if } V(A)(w) = 0 \end{cases}$$

Now with $V(A \rightarrow C)$ defined everywhere, the desired equality of the probability of the conditional with the conditional probability follows as shown in (25).

(25) $$P(A \rightarrow C) = E[V(A \rightarrow C)]$$
$$= E[V(A \rightarrow C)|\{V(A) = 1\}] \cdot \left(Pr(\{V(A) = 1\}) + Pr(\{V(A) = 0\}) \right)$$
$$= Pr(\{V(C) = 1\}|\{V(A) = 1\})$$
$$= P(C|A)$$

So far, so good. What remains to be seen is whether this "rather weird three-valued entity" (Edgington, 1995) deserves to be credited with any explanatory value. As it stands, it appears to be little more than technical "hack" whose utility as a theory of conditionals is rather questionable. In particular, it invites two objections: Firstly, as it is defined, V is really a function of *three* arguments: A sentence, a world, and the probability distribution Pr. The value at one and the same world will change if the probability distribution changes. In what sense, then, is that value "objective"? Secondly, what does it *mean* for values at individual worlds—in the absence of any uncertainty about the facts—to be intermediate? Why should it be sensible to assign such values? I will address these points in the following section.

4.2. *Time, truth, and chance*

At worlds at which the antecedent is false, the value of a conditional depends on the probability distribution Pr. Does this amount to abandoning the notion that the value is *objective*, i.e., that it only depends on the *facts* at the world of evaluation? Not if the distribution Pr itself only depends on the facts at that world. It may not be immediately clear how that should be, however, for probabilities arise under uncertainty, but a fully specified world admits of no uncertainty.

The key to reconciling these apparently contradictory notions is the fact that at the *time* the conditional is evaluated, the world is *not yet* fully specified—only an initial segment of its history is. This initial segment also determines the set of its possible continuations and their respective probabilities: the (objective) *chance* that the *actual* continuation is of one kind rather than another.

Objective chances manifest themselves in the distribution of facts. Put simply, chances are the predictions made by a *good* stochastic model which fits the stream of facts *well*. More can be said about the goodness of a model and the goodness of its fit, but that is not a linguistic problem. For discussions, see for instance Ramsey (1929); Lewis (1980, 1994); Grünwald (1998).

A model in which the evolution of chance over time can be encoded must add a temporal dimension to the set of worlds. The Venn diagrams in Section 3.3 depict worlds as points. If we "pull them out of the page and lay them on the side," they look like Figure 3. Models of this kind have been dubbed "$T \times W$-frames" by Thomason (1984). Definition 2 is his.

Definition 2 (T $\times$ W-frames (Thomason, 1984)) *A $T \times W$ frame is a quadruple $\langle W, T, <, \approx \rangle$, where (i.) W and T are nonempty sets, (ii.) $<$ is a transitive relation on T which is also irreflexive and linear (i.e. $t \not< t$ for all $t \in T$, and either $t < t'$ or $t' < t$ or $t = t'$ for all $t, t' \in T$), and (iii.) $\approx$ is a 3-place relation on $T \times W \times W$, such that*

a. *for all t, $\approx_t$ is an equivalence relation (i.e. $w \approx_t w$ for all $t \in T$ and $w \in W$, etc.), and*

b. *for all $w_1, w_2 \in W$ and $t, t' \in T$, if $w_1 \approx_t w_2$ and $t' < t$, then $w_1 \approx_{t'} w_2$.*
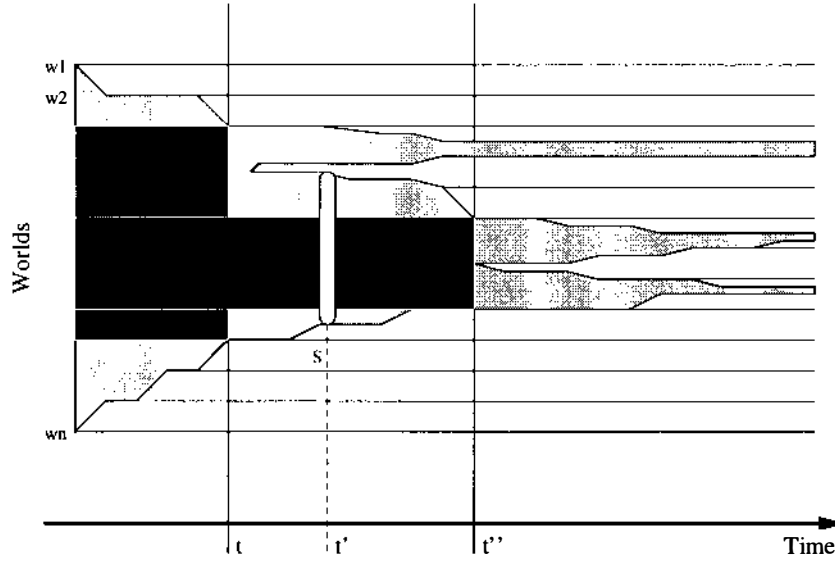
Figure 3: A $T \times W$-model.

The conditions imposed by Definition 2 enable the model to encode some fundamental intuitions about time. (ii.) requires that points in time be ordered as shown in Figure 3. The purpose of the $\approx$-relation in (iii.) is to define the "historical alternatives" to a given world w at a time t as those worlds which agree with w at all past times through t but may differ in the future. This accounts for the intuition that different futures can grow out of identical (or in any case, indistinguishable) pasts.

Our world evolves through time by shedding alternative courses of history. At each time, exactly one of the states it *could have* been in is the one it actually *is* in. Furthermore, at each time, past and present are fixed: Although they *could have* been otherwise, they *cannot* (any longer) have been otherwise. This is known as *historical necessity*.

It will be helpful the reify the sets of worlds which agree with the *actual* course of history up to some given time.

Definition 3 (Histories in T × W-frames) *Let* $[w]_t = \{w' \in W \mid w \approx_t w'\}$ *be the set of historical alternatives of w at t. A* history *is a function* $h : T \mapsto \wp(W)$ *such that for some world* $w \in W$, $h(t) = [w]_t$ *at all* $t \in T$.

Clause (iiib.) on the $\approx$-relation imposes a *monotonicity* condition on histories: Each $h(t)$ is a subset of $h(t')$ for all $t' < t$. Thus it is impossible for two worlds to "become" historical alternatives at some point. Nor is it possible for two historical alternatives "not to have been" historical alternatives at earlier times.

Next, the assignment function V is made sensitive to time. Let $\mathcal{A}$ bet he set of atomic sentences in the language.

Definition 4 (T × W interpretation) *A* valuation function *is a function* $V : \mathcal{A} \times T \times W \mapsto \{0, 1\}$ *from atomic sentences and time-world pairs to truth values, provided that for all atomic sentences* $A \in \mathcal{A}$, *if* $w \approx_t w'$, *then* $V(\varphi)(t)(w) = V(\varphi)(t)(w')$.

This extends to truth-conditional compounds straightforwardly, as before. Definition 4 ensures that the worlds belonging to the "slice" $h(t)$ of history h at time t agree on all facts at all times leading up to t. In Figure 3, this is visualized by the black rectangles: Worlds that are enclosed by a black rectangle are indistinguishable at all times covered by the black area.

Finally, in dealing with predictions, a means of representing and evaluating assertions about the future is indispensable. In addition to the quantifiers P and F familiar from Prior's (1967) tense logic (and their duals H and G,) it is also convenient to have expressions of the form φ^t with a temporal index, paraphraseable as 'φ *was / is / will be true at time t*'. The valuation function is extended as expected:

(26) $V(\varphi^{t'})(t)(w) = V(\varphi)(t')(w)$

(27) $V(P\varphi)(t)(w) = \begin{cases} 1 \text{ if for some } t' \text{ s.t. } t' < t, V(\varphi)(t')(w) = 1 \\ 0 \text{ otherwise} \end{cases}$

(28) $V(F\varphi)(t)(w) = \begin{cases} 1 \text{ if for some } t' \text{ s.t. } t < t', V(\varphi)(t')(w) = 1 \\ 0 \text{ otherwise} \end{cases}$

Definition 4 combined with this extension to tensed sentences implies that all worlds belonging to a "slice" $h(t)$ of history agree not only on the truth values of all sentences at all times leading up to t, but also on the the truth values of all sentences *about* times up to t. No such requirement follows for statements about times later than t, however.

This leads to the last notion to be introduced here, and the one most relevant in the present context: The truth of a sentence is *settled* in a history h at a time t whenever it is *impossible*, relative to the set $h(t)$, for it to be false. Again following Prior (1967), I define the operator L; $L\varphi$ stands for 'φ *is settled.*'

(29) $V(L\varphi)(t)(w) = \begin{cases} 1 \text{ if for all } w' \in h(t), V(\varphi)(t)(w') = 1 \\ 0 \text{ otherwise} \end{cases}$

The distinction between truth and settledness is important and bears some potential for confusion. The two are equivalent for statements not about the future: Whatever is *true* about past and present is also *settled*. The future, in contrast, is *open*: Statements about it may be *true* without being *settled*. It is possible to make true predictions without any way of knowing (already) that they are true.

Our readiness to assert or assent to predictions is guided by the settledness, so that the intended notion of truth requires some getting used to. We are most confident in those predictions whose truth is established. However, even unlikely predictions are true as long as they "turn out" true in due course. Consider a sentence

$\varphi^{t'}$ at time t, where $t < t'$. As time progresses beyond t' and continuations that were possible at t are eliminated, only those worlds w survive for which $V(\varphi^{t'})(t)(w)$ *was* 1 or 0, as the case may be, *already*. The truth value of $\varphi^{t'}$ at t is established "retroactively," but its truth does not imply settledness. Thus simple predictions like (30) have truth values at all worlds.

(30)   You will strike the match.

It is important to emphasize here the difference between sentences like (30) and conditionals, which as I claimed above do not always have truth values. This claim does *not* mean that their truth is merely not settled in the way the truth of simple predictions is not settled. Rather, it means that they differ more fundamentally from simple predictions like (30) in that at some worlds their truth cannot be established at all.

The final step is the re-introduction of probability, now interpreted as objective chance. In Section 3.3, Pr was defined over a set of timeless worlds, modeled as points. Now the points are stretched out into world-lines, but no additional apparatus is needed to model the evolution of chance over time. Pr is the objective *prior* distribution over the possible courses of history. This is an idealization from a metaphysical point of view—one may not believe that chance exists in the absence of facts—but one we are free to make in building the model. As time progresses and the facts "come in," the set of worlds among which ours must be becomes smaller, and the probability mass is accordingly concentrated on ever smaller subsets of W. New distributions $Pr_{h,t}$ derive from Pr by *conditioning* on the totality of settled facts embodied by $h(t)$.

Definition 5 (Probability in history) *Given a $T \times W$-frame and a probability distribution Pr over W, the distribution $Pr_{h,t}$ in history h at time t is defined as follows: For all propositions X, $Pr_{h,t}(X) = Pr(X|h(t))$.*

Definition 5 makes a connection between settledness and chance and populates the space between 0 and 1 with intermediate degrees. The settledness of past and present facts corresponds to extreme chances: The chance that the match lit is 1 if it did, 0 if it did not. The chance that the match *will light*, on the other hand, is intermediate unless its lighting is either inevitable or impossible. Probabilities of *sentences* at times are obtained as before via V and Pr, as expectations conditional upon the history h to t:

(31) $P^t(\varphi^{t'}) = E[V(\varphi^{t'})(t)|h(t)] = E[V(\varphi)(t')|h(t)]$

(*) $= Pr(\{V(\varphi)(t') = 1\}|h(t))$

If φ is atomic or truth-functional, the value comes down to (*); conditionals, however, receive intermediate values where their antecedents are false, as outlined in Section 3.5.

The first potential objection to the intermediate-value approach is now addressed: The probability distribution with respect to which the intermediate values at worlds are calculated, interpreted as objective chance, is fully determined by the facts at the world and time of evaluation (recall that all worlds in $h(t)$ are indistinguishable with respect to their past.)

The second potential objection concerned the very idea of assigning intermediate values to conditionals at worlds. Truth *values* are defined by truth *conditions* with reference to *facts*. Simple predictions about the future depend for their truth values on future facts. Are not the truth values of conditionals established "retroactively" as well? Consider (32a), uttered at some time, say 9a.m., about your behavior over the next couple of minutes.

(32) a. If you strike the match, it will light.
      b. You will strike the match.

To avoid the potentially confusing talk of "truth," it may be safer to speak of the *verifying* and *falsifying* instances among the worlds in $h(t)$. What future facts reveal about the simple prediction in (32b) is not whether it was *settled*, but whether it *turns out true*—that is, whether the actual world is one of the verifying cases. Likewise for (32a). What do future facts at a later time, say 10a.m., reveal about (32a)?

Those worlds in which you do in fact strike the match are easy to classify: Where it lights, (32a) is *verified*; where it doesn't, (32a) is *falsified*. Things are much less clear where you don't strike the match. Whether it lights (for some other reason) or not has no bearing on the truth of the conditional. Classical logic stipulates counterintuitively that your not striking the match is sufficient for the *truth* of (32a). It would be no more intuitive to stipulate that (32a) is *false* at those worlds. Nor should its value be *undefined*: Your not striking the match is does not mean that we know *less* about it then before. Indeed, (32a) is just "as true" as it was when it was used. This intuition is captured in the assignment of intermediate values at worlds where the antecedent is false.

4.3. *Other conditionals*

Although this paper is mainly concerned with the objective values of predictive conditionals, a brief remark on the values of other conditionals is in order. Compare (33a), used at 10a.m., when the facts are settled, with (32a).

(33) a. If you struck the match, it lit.
      b. If you had struck the match, it would have lit.

The facts relevant to the truth of (33a) are settled; there is no *objective* uncertainty. At worlds at which you did not strike the match, it is impossible for you to have struck it. The objective value of (33a) at 10a.m. is undefined. Accordingly,

the use of (33a) by a speaker who *knows* that you did not strike the match is infelicitous. However, (33a) may be used felicitously by a speaker who is *ignorant* of the relevant facts.[7]

Thus the values of (33a) at 10a.m. and (32a) at 9a.m. are quite different. Not so with (33b): Intuitively, its value at 10a.m. should be the same as that of (32a) at 9a.m. This intuition is on the right track, although it needs to be refined in ways that need not concern us here (cf. Kaufmann (2001).) The connection between predictions and counterfactuals is helpful in the next section, however.

5. Right-nested conditionals

The assignment function V as defined above extends without further modification to conditionals whose apodosis is itself a conditional. It turns out, however, that the values predicted for such complex conditionals are counterintuitive. In this last section I am going to discuss one example of this kind to motivate an amendment to the truth definitions. There is not enough room to discuss more embedded conditionals and compounds thereof. Instead, the discussion is intended to illustrate the connection between predictions and counterfactuals in some more detail. In the interest of readability I will return to the simpler definition of V without reference to time.

5.1. *Wrong predictions*

Consider the right-nested conditional in (34a), symbolized as (34b).

(34)   a.   If the match is wet, then if you strike it, it will light.
        b. $W \rightarrow (S \rightarrow L)$

To test the predictions, suppose the circumstances are as summarized in (35). Intuitions are not generally as clear-cut as the customary use of real numbers would suggest. Accordingly, the numbers in (35) are merely intended as formal stand-ins for rough estimates such as "high," "low," etc.

(35)   a.   Whether you strike the match is independent of whether it is wet.
        b.   The probability that ...
             – it gets wet is "low" (.2)
             – you strike it is .5
             – it lights if it is dry and you strike it is "high" (.8)
             – it lights if it is wet and you strike it is "tiny" (.01)

From (35) it follows that $P(L|S)$, the probability that it lights given that you strike it, is $P(L|SW)P(W|S) + P(L|S\overline{W})P(\overline{W}|S) = .01 \cdot .2 + .8 \cdot .8 = .642$.

Figure 4a displays the distribution of values of the consequent $S \rightarrow L$ of (34.) It is 1 (black) where both S and L are true, 0 (white) where S is true and L is

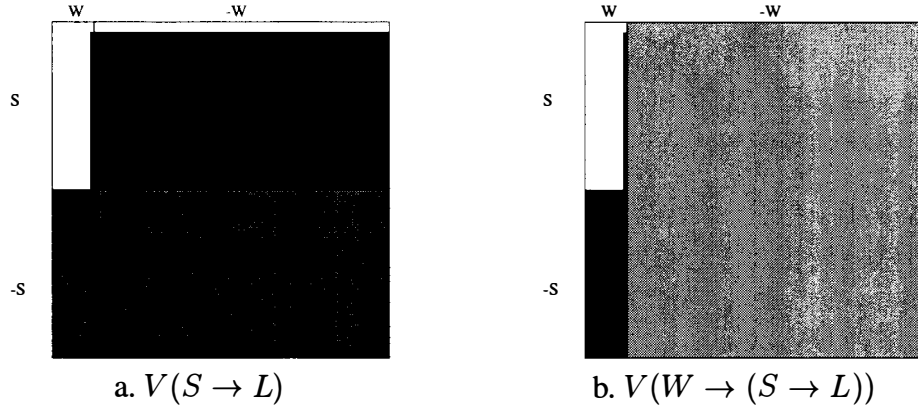a. $V(S \to L)$ b. $V(W \to (S \to L))$

Figure 4: Distribution of values for (34)

false, and intermediate (grey) where S is false. The relative sizes of the areas in the figure roughly correspond to the probabilities given in (35).

Intuitively in a situation like this, the probability of (34) should be "tiny." That is not, however, what the definitions predict. The assignment of values of (34) is defined as in (36).

(36) $V(W \to (S \to L))(w)$
$$= \begin{cases} V(S \to L)(w) \text{ if } V(W)(w) = 1 \\ E[V(W \to (S \to L))|\{V(W) = 1\}] \text{ if } V(W)(w) = 0 \end{cases}$$

where the values of $V(S \to L)$ are

(37) $V(S \to L)(w)$
$$= \begin{cases} V(L)(w) \text{ if } V(S)(w) = 1 \\ E[V(S \to L)|\{V(S) = 1\}] \text{ if } V(S)(w) = 0 \end{cases}$$

To calculate $P(W \to (S \to L))$, the expectation of the values of (36), we first consider the values at those worlds where its antecedent is true, i.e., where the match is wet. In Figure 4 this corresponds to the two quadrants on the left. At all those worlds, the value of the consequent is assigned as the value of the conditional.

(38) $P(S \to L|W) = E[V(S \to L)|\{V(W) = 1\}]$
$$= 1 \cdot Pr(\{V(S) = V(L) = 1\}|\{V(W) = 1\})$$
$$+ 0 \cdot Pr(\{V(S) = 1, V(L) = 0\}|\{V(W) = 1\})$$
$$+ .81 \cdot Pr(\{V(S) = 0\}|\{V(W) = 1\})$$
$$= 1 \cdot .01 + .642 \cdot .5$$
$$= .331$$

Again following the overall strategy, this expectation of the conditional given that its antecedent is true (i.e., that the match is wet) is spread uniformly over those worlds where it is false (where the match is dry.) This distribution is illustrated in Figure 4b. As a result, the overall probability $P(W \rightarrow (S \rightarrow L))$ is .331 as well. But this is certainly not the "tiny" probability expected in the given scenario.

5.2. *Fixing the rule*

What went wrong in the preceding example? The connection between the values of predictive conditionals and counterfactuals, pointed out above, is particularly helpful here. It is possible to have intuitions about the source of the problem, and those intuitions suggest a fairly straightforward repair.

Suppose we inhabit a world w in which you did not strike the match. Recall that facts like this determine the value (39a) had at w at some earlier time. What was that value?

(39) a. If you strike the match, it will light.
      b. If you had struck the match, it would have lit.

There is no clear answer to this question. We have not been told enough about w: The chance depends on a "third fact," *viz.* whether the match is wet or not, in much the same way as the value of (39b) does.

In section 4.2 I argued that the values of conditionals at worlds at which their antecedents are false are the *objective chances* of the consequent, given that the antecedent is true. Objective chance varies depending on many factors. The condition of the match affects the value of the conditional. The relation of "affecting the chance of" is typically part of the definition of *causal dependencies* (cf. Davis (1988) for an overview.) Without complicating the formalism with further technicalities, we may just note two relevant facts about the scenario:

1. Whether the match lights causally depends on both whether it is wet and whether you strike it.

2. Whether you strike the match and whether it is wet are causally independent.

Informally, the import of these causal dependencies is the following: Given the past facts up to the time in question, the process that brought it about that you did not strike the match is independent of the process that brought about the wetness of the match. Therefore at a world where the match is wet and you did not strike it, the chance that it would have lit if you had struck it is the chance that it would have done so *given that it is wet.*[8]

Thus although the counterintuitive predictions of the definitions above are only revealed in complex sentences such as (34), their source lies in the assignments to simple conditionals. In Figure 4a, the values of $V(S \rightarrow L)$ are uniform on the set of worlds at which the match is not struck. This is, in fact, made explicit, though not explained, by Jeffrey (1991):
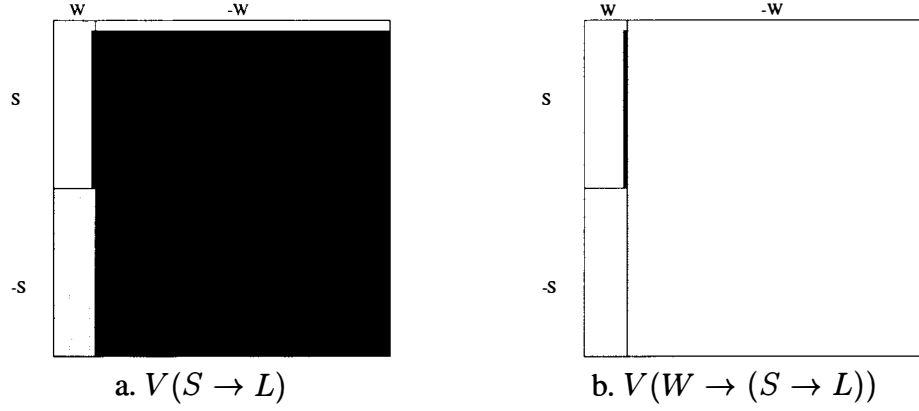
a. $V(S \to L)$ b. $V(W \to (S \to L))$

Figure 5: Distribution of values for (34) accounting for causal relations

> " ... if $A \to C$ is to be an *indicative* conditional it must have the same value at all worlds where A is false." (p. 172)

It is not clear why this should be so, and examples like (34) suggest that it is wrong. Figure 5a shows an alternative assignment, accounting for causal dependencies: The values assigned at the worlds where the match is not struck are not uniform, but vary depending on the condition of the match. Formally, this assignment is defined as in (40).

(40) $V(A \to C)(w) =$
$$\begin{cases} V(C)(w) \text{ if } V(A)(w) = 1 \\ E[V(A \to C)|\{V(\varphi_i) = V(\varphi_i)(w)\}] \text{ if } V(A)(w) = 0 \\ \qquad \text{for all } \textit{non-descendants } X_i \text{ of } A \text{ in the causal order.} \end{cases}$$

The probabilistic value of (34) then is indeed the "tiny" value that was expected, given the scenario. (41) derives its expectation given that the match is wet; according to the rule, that value is then assigned to all worlds at which the match is not wet, as shown in Figure 5b. Hence the overall expectation is "tiny" indeed:

(41) $P(W \to (S \to L)|\{V(W) = 1\})$
$$\begin{aligned} &= 1 \cdot Pr(\{V(S) = V(L) = 1\}|\{V(W) = 1\}) \\ &\quad + 0 \cdot Pr(\{V(S) = 1, V(L) = 0\}|\{V(W) = 1\}) \\ &\quad + E[V(S \to L)|\{V(S) = V(W) = 1\}] \\ &\qquad\qquad \cdot Pr(\{V(S) = 0\}|\{V(W) = 1\}) \\ &= .005 + .01 \cdot .5 \\ &= .01 \end{aligned}$$

6. Conclusions

This paper discussed the objective values of predictive conditionals. Taking those values to correspond to "degrees of truth," a pragmatic question arises: When is a sentence, conditional or otherwise, "true enough" to be used in inference or conversation? The probabilistic explanation of the patterns in Section 2 rests on the assumption that premises whose probability is short of 1 are admissible as long as their probabilities are "close enough" to 1. How close is close enough?

I believe that no definitive answer can be given to this. However, the lack of a clear cut-off point need not be thought of as a deficiency and can be explained away in much the same way as is done in the treatment of *vagueness*: by assuming, at least for the sake of argument, that in each context of use an implicit parameter is given which decides how true is "true enough." In Adams' definition of probabilistic entailment (Endnote 2) this role is played by the variable ϵ: Roughly speaking, an inference is valid if its consequence can be made "true enough" under any interpretation of "true enough."

A probabilistic semantics promises to be useful on three levels: Assuming, as I do, that contemporary science is right and the processes governing the world are truly non-deterministic, the objective values of predictive conditionals are as discussed above. Short of such metaphysical commitments, even if the world is deterministic, the fact remains that we talk "as if" it were not; otherwise coin tosses and other phenomena would not make for such convincing illustrations of random processes. And it is the way we talk that is at issue here. Finally, even if predictions could in principle be made with certainty, a theory of language interpretation which accommodates the fact that we cannot and need not perform such calculations—one of approximate inference—is still desirable.

Uncertainty arises in language use in a number of ways, only two of which are ignorance and physical non-determinism. Other sources, such as vagueness, lexical and syntactic ambiguity, are instances of *linguistic uncertainty* with semantic consequences. A large array of topics related to the one of this paper, such as the relationship between uncertainty about word denotations and uncertainty about sentence denotations, are only beginning to emerge.

Endnotes

1.   Throughout this paper I follow the following notational conventions: The material conditional is symbolized by '$\supset$,' negation by a horizontal bar; thus $A \supset C \equiv \overline{A} \vee C$. The "natural" conditional corresponding to English '*If* ... *then* ... ' is referred to by '$\rightarrow$.' Linguistic examples with labels of the form [nyt-.-] are attested in the New York Times corpus; the numbers refer to their location in the corpus.

2.   More precisely, Adams' definition of "p-validity" is as follows: "A sentence φ follows from a set of sentences Γ if and only if 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 $B \in \Gamma$, $Pr(\varphi) \geq 1 - \epsilon$."

3.   The third condition must in general hold for *countable* unions. This complication is irrelevant as long as we deal with (finite sets of) finite sentences.

4.   In general, integrals would replace the sums.

5.   Kaufmann (2001) discusses Lewis' triviality results in some more detail.

6.   In statistical jargon, the function $V(A \to C)$ denoted by a conditional is a *random variable*. The two-valued function $V(A)$ denoted by an atomic sentence A is an *indicator function*. The latter are special cases of the former.

7.   Ignorance of settled facts can be modeled as a subjective distribution over multiple possible histories (ways of "filling in the blanks") only one of which can be the actual one. Subjective probabilities are estimates of the corresponding objective chances. A speaker may entertain the possibility that a conditional probability is defined when in fact it is not.

8.   Technically, causal dependencies of the kind discussed here form a strict partial order on the facts of the world, in which the wetness of the match is not a descendant of the striking. All non-descendants of the antecedent are "held constant" when examining alternative worlds in which the antecedent is true. This is closely related to the treatment of "structural counterfactuals" by Pearl (2000).

References

Adams, E. 1965. The logic of conditionals. *Inquiry*, 8:166–197.

Adams, E. 1975. *The Logic of Conditionals*. Reidel.

Adams, E. 1998. *A Primer of Probability Logic*. CSLI Publications.

The Aristotelian Society. 1991. *The Symposia Read at the Joint Session of the Aristotelian Society and the Mind Association at the University of Durham.* Supplementary Volume 65.

Davis, W. A. 1988. Probabilistic theories of causation. In Fetzer, J. H., editor, *Probability and Causality*, volume 192 of *Studies in Epistemology, Logic, Methodology, and Philosophy of Science*, pages 133–160. D. Reidel, Dordrecht / Boston / Lancaster / Tokyo.

Edgington, D. 1995. On conditionals. *Mind*, 104(414):235–329.

van Fraassen, B. C. 1976. Probabilities of conditionals. In Harper, W. L., R. Stalnaker, and G. Pearce, editors, *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, volume 1 of *The University of Western Ontario Series in Philosophy of Science*, pages 261–308. D. Reidel.

Gabbay, D. and F. Guenthner, editors. 1984. *Extensions of Classical Logic*, volume 2 of *Handbook of Philosophical Logic*. D. Reidel.

Grünwald, P. 1998. *The Minimum Description Length Principle and Reasoning under Uncertainty*. PhD thesis, ILLC, Amsterdam.

Harper, W., R. Stalnaker, and G. Pearce, editors. 1981. *Ifs: Conditionals, Belief,*

Decision, Chance, and Time. Reidel.

Jackson, F., editor. 1991. *Conditionals*. Oxford University Press.

Jeffrey, R. C., editor. 1980. *Studies in Inductive Logic and Probability*, volume 2. University of California Press.

Jeffrey, R. C. 1991. Matter-of-fact conditionals. In *The Symposia Read at the Joint Session of the Aristotelian Society and the Mind Association at the University of Durham*, pages 161–183. The Aristotelian Society. Supplementary Volume 65.

Kaufmann, S. 2001. *Aspects of the Meaning and Use of Conditionals*. PhD thesis, Stanford University.

Kratzer, A. 1979. Conditional necessity and possibility. In Egli, U., R. Bäuerle, and A. von Stechow, editors, *Semantics from Different Points of View*, pages 117–147. Springer.

Kratzer, A. 1986. Conditionals. In Farley, A. M., P. T. Farley, and K.-E. McCullough, editors, *Proceedings of CLS 22, Part 2: Papers from the Parasession on Pragmatics and Grammatical Theory*, pages 1–15.

Kratzer, A. 1991. Conditionals. In von Stechow, A. and D. Wunderlich, editors, *Semantik: Ein internationales Handbuch der zeitgenössischen Forschung. [=Semantics]*, pages 651–656. de Gruyter.

Lewis, D. 1973. *Counterfactuals*. Harvard University Press.

Lewis, D. 1976. Probabilities of conditionals and conditional probabilities. *Philosophical Review*, 85:297–315. Reprinted with a Postscript in (Jackson, 1991, 76–101).

Lewis, D. 1980. A subjectivist's guide to objective chance. In Jeffrey (1980), pages 263–293. Reprinted in (Harper et al., 1981, 267–297) and with a Postscript in (Lewis, 1986a, 83–132).

Lewis, D. 1986a. *Philosophical Papers*, volume 2. Oxford University Press.

Lewis, D. 1986b. Probabilities of conditionals and conditional probabilities II. *Philosophical Review*, 95:581–589. Reprinted in (Jackson, 1991, 102–110) and (Lewis, 1998, 57–65).

Lewis, D. 1994. Humean Supervenience debugged. *Mind*, 103:473–490. Reprinted in (Lewis, 1999, 224–247).

Lewis, D. 1998. *Papers in Philosophical Logic*. Cambridge University Press.

Lewis, D. 1999. *Papers in Metaphysics and Epistemology*. Cambridge University Press, Cambridge.

Mellor, D., editor. 1990. *Philosophical Papers: F.P. Ramsey*. Cambridge University Press, Cambridge, U.K.

Pearl, J. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge University Press.

Prior, A. 1967. *Past, Present and Future*. Oxford University Press.

Ramsey, F. 1929. General propositions and causality. In Mellor (1990), pages 145–163.

Thomason, R. H. 1984. Combinations of tense and modality. In Gabbay and Guenthner (1984), pages 135–165.