Local and Global Interpretations of Conditionals

Stefan Kaufmann
Department of Linguistics
Northwestern University
kaufmann@northwestern.edu

September 11, 2005

Abstract

The unified probabilistic theory of indicative and counterfactual conditionals proposed by Kaufmann (2005a) leads to apparently conflicting predictions about the probabilities of indicatives. We present experimental data which show that these predictions are not only not at odds with the theory, but in fact reveal a real and rarely discussed ambiguity.

1 Introduction

The thesis that the subjective probability of a conditional *'if A, C'* is the conditional probability of C, given A (henceforth "the Thesis") has a long history in the philosophical literature, starting with Ramsey's (1929) suggestion that conditionals are used to argue about "degrees of belief" in C given A. Philosophers have since extensively studied itsf ramifications for logical and semantic theory (see Eells and Skyrms, 1994; Edgington, 1995; Bennett, 2003 for recent overviews of this extensive literature). More recently, psychologists have begun to recognize its value in explaining subjects' behavior in a variety of experimental settings without the complexities and additional assumptions involved in alternative accounts, such as the theory of mental models (see Oaksford and Chater, 2003; Over and Evans, 2003 and references therein).

Despite its simplicity and intuitive appeal, the Thesis raises important theoretical and empirical questions which remain yet to be resolved. Three such open issues form the background of the study reported in this paper.

The first is the technical problem of integrating the Thesis with a propositional logical theory of the familiar kind. Lewis' (1976; 1986) famous *triviality results*, and the subsequent investigations they inspired, show that the Thesis is incompatible with the assumption that conditionals denote propositions in the usual sense. The second question concerns the relationship between (predictive) *indicative* conditionals like (1a) and their *counterfactual* counterparts (1c). Many proponents of the Thesis simultaneously embrace some version of the claim that the (posterior) probability of a counterfactual, at the time at which its use is appropriate, is the (prior) probability of its indicative counterpart at an earlier time, and that this relationship extends in some straightforward way to cases in which the relevant notions of "prior" and "posterior" are not temporal (Edgington, 1995). Barker (1998) dubbed this assumption "Tense Probabilism" and argued convicingly that despite its plausibility in particular cases, it does not hold as a general principle.

(1) a. If she throws an even number, it will be a six.
    b. If she threw an even number, it was a six.
    c. If she had thrown an even number, it would have been a six.

The third problem is that the Thesis does not even accord well with intuititions for all indicative conditionals, as is shown by counterexamples that have occasionally been discussed in the philosophical literature (Pollock, 1976; McGee, 2000, among others). Examples of this kind are at the center of the study reported below. There is no consensus on the question of how to reconcile such diverging intuitions with the intuitive correctness of the Thesis in most other cases.

A unified probabilistic theory of conditionals must give answers to these questions. Recently, Kaufmann (2005a) offered a proposal concerning the relationship between indicative and counterfactual conditionals. As I will discuss in some detail below, this proposal gives rise to apparently contradictory predictions about the probabilities speakers will assign to indicative conditionals in certain circumstances. Kaufmann (2004) discusses some relevant examples and argues that far from being problematic, these conflicting predictions in fact point to a rarely discussed indeterminacy and context dependence in the interpretation of conditionals.

This paper presents the results of a pilot study designed to test the predictions of Kaufmann's theory. Subjects were asked to judge the probabilities of conditionals based on incomplete information about the relevant facts. The facts were presented in scenarios of the kind for which Kaufmann's theory predicts probability judgments to vary. The results of this preliminary study largely confirm Kaufmann's claims. Not only did subjects' judgments exhibit the predicted variation, they were also influenced by the linguistic context in which stimuli were presented as well as the presententation of the facts, in ways that are consistent with Kaufmann's analysis of the inferences involved in evaluating conditionals.

I will begin by outlining the theoretical considerations that motivated Kaufmann's account in Section 2. Section 2.6 presents the problematic predictions of the account and Kaufmann's defense. These discussions will be brief; the reader is referred to the works cited for further details and references. Section 3 discusses the experiments and their theoretical import. Section 4 concludes by pointing out further open questions and directions for future work.

2 Background

While Kaufmann's theory is insprired by the Thesis (that the probabilities of conditionals are conditional probabilities), it adapts and partially departs from it in two major ways. First, in incorporating the Thesis in standard possible-worlds semantics, a non-standard way of assigning truth values to conditionals ensures that the resulting system is not subject to Lewis' triviality results. Second, these non-standard truth values are made sensitive to causal dependencies in order to yield intuitively correct predictions about counterfactual conditionals. The result is intended as a unified account of both indicative and counterfactual conditionals. In this section I discuss its main building blocks in some more detail.

2.1 The Ramsey Test

Like most semantic theories of indicative conditionals, the probabilistic approach appeals to an intuition that was first spelled out by Ramsey (1929):[1]

> If two people are arguing 'If A will C?' and are both in doubt as to A, they are *adding A hypothetically* to their *stock of knowledge* and arguing on that basis about C... We can say they are fixing their *degrees of belief* in C given A.

In order to make this informal suggestion precise, the notions emphasized above must be made precise. In a probabilistic framework, the following assumptions are the most straightforward way to do so:

[1] Emphasis added. For consistency, I replace Ramsey's p and q with A and C, respectively.

(2) a. "Stocks of knowledge" are represented by sets of possible worlds.
b. "Degrees of belief" correspond to (subjective) probability distributions over possible worlds.
c. The "addition of A" to a stock of knowledge proceeds by conditionalization.

I assume familiarity with possible worlds, the framework in which virtually all treatments of conditionals, probabilistic or otherwise, are couched. A probability distribution over a set W is a function Pr from subsets of W to the interval $[0, 1]$ such that (i) $Pr(W) = 1$ and (ii) for all disjoint subsets X, Y of W, $Pr(X \cup Y) = Pr(X) + Pr(Y)$.[2]

The probability distribution Pr is defined for propositions (i.e., sets of worlds), but it is the sentences of the language that are of interest here. To make this connection, atomic sentences are assigned truth values point-wise at possible worlds by a valuation function V. This assignment extends to truth-functional compounds of sentences as dictated by the rules of standard propositional logic:

(3)
$$\begin{aligned} V_w(\neg\varphi) =& 1 - V_w(\varphi) \\ V_w(\varphi \wedge \psi) =& V_w(\varphi) \cdot V_w(\psi) \end{aligned}$$

Other connectives can be defined in terms of those in (3). Now in statistical terms, the denotation $V(\varphi)$ of a sentence φ is a *random variable* on the probability space $\langle W, Pr \rangle$. Based on V and Pr, a probability distribution P over sentences is defined: The probability $P(\varphi)$ of a sentence φ is the *expectation* of its truth value. The expectation may be written '$E[V(\varphi)]$' and is defined as the weighted sum of the values of $V(\varphi)$, where the weights are the probabilities that $V(\varphi)$ has those values:[3]

(4)
$$P(\varphi) = E[V(\varphi)] = \sum_{x \in \{0,1\}} x \cdot Pr(V(\varphi) = x)$$

For the sentences discussed so far, V ranges over the set $\{0, 1\}$ of truth values, thus the probability of a sentence is the probability that it is true.

[2] The second requirement must also hold for the limits of countable unions, but I ignore this here for simplicity.

[3] I write '$Pr(V(\varphi) = x)$' instead of '$Pr(\{w \in W | V_w(\varphi) = 1\})$' for the probability of the event that $V(\varphi)$ has value x.

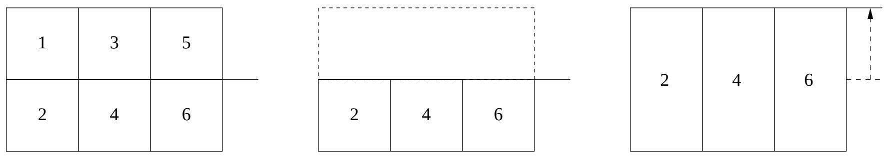

Figure 1: Update with A in two steps: elimination and renormalization

Thus for instance, the probability of *'she throws an even number'* is just the probability that she throws an even number.[4]

Ramsey suggests that a conditional *'if A then C'* should be analyzed in terms of the dynamics of belief change — specifically, the result of an update of the probability distribution Pr with the information that the antecedent A is true. This update plays an important role below and deserves some discussion at this point. It is useful to think of it as an operation involving two steps, *elimination* and *renormalization*. The former removes those possible worlds at which A is false (by setting the probability of the corresponding proposition to 0). The latter consists in recalibrating the probabilities of the remaining sets of worlds (those at which A is true). The updates proceeds by conditionalization if this renormalization is carried out in such a way that the relative probabilities of all remaining propositions with non-zero probability are preserved.

For concreteness, consider again the conditional in (1a), repeated here as (5):

(5) If she throws an even number, it will be six.

The two-step update may be visualized as in Figure 1. The numbered cells represent the six equally likely outcomes of the toss of a fair die. After the elimination of those outcomes at which the antecedent is false (i.e., the number is odd), the probabilities are recalculated so that the antecedent receives probability 1 and the proportions within the remaining set of worlds are restored. The third picture shows the result. The probability that the number is six has grown from 1/6 in the original distribution to 1/3. This is the *posterior* probability of the consequent. The Thesis states that this posterior probability equals the *prior* probability of the conditional (5).

[4] With the proper distinction between object language and metalanguage, this statement is no more circular than the Tarskian slogan that *'snow is white'* is true if and only if snow is white.

2.2 Truth values for conditionals

Stated in the terms of this framework, the most intensely debated philosophical problem with the Thesis is that of extending the assignment function V to conditionals. Just like for other sentences, the expectation of the value $V(\text{if } A \text{ then } C)$ should be its probability — that is, the conditional probability of C, given A. This goal has proven elusive. A conditional probability cannot in general be interpreted as the probability that a proposition is true. Lewis (1976, 1986) showed that except for certain trivial cases, it is impossible to assign a proposition to a conditional in such a way that its probability is guaranteed to equal the corresponding conditional probability for any probability distribution Pr.[5]

One *can*, however, assign values to the conditional which depend on Pr. The following definition was proposed by Jeffrey (1991), later elaborated by Stalnaker and Jeffrey (1994), and adopted with certain modifications by Kaufmann (2005a):

(6) $$V_w(\text{if } A \text{ then } C) = \begin{cases} V_w(C) \text{ if } V_w(A) = 1 \\ E[V(C)|V(A) = 1] \text{ if } V_w(A) = 0 \end{cases}$$

Thus at worlds at which A is true, the conditional is equivalent to its consequent (hence to the material conditional). At worlds at which A is false, the value of the conditional is the conditional expectation of $V(C)$, given that A is true. This value equals the conditional probability of C, given A, and may fall anywhere in the interval $[0, 1]$.

Under this definition, the conditional in (5) receives three different values:

(7) $$V_w(5) = \begin{cases} 1 \text{ if the number is even and six at } w \\ 0 \text{ if the number is even and not six at } w \\ 1/3 \text{ if the number is odd at } w \end{cases}$$

The resulting distribution is shown in Figure 2. Shades of grey indicate the values: black for 1, white for 0, and grey for 1/3.

This strategy avoids Lewis' triviality results, but it does so in virtue of two rather unconventional features: The value of a conditional at a non-

[5] A detailed discussion of this problem is beyond the scope of this paper. Lewis' papers are still the best expositions of the original results. Subsequent authors extended and generalized them in various ways. See Hájek and Hall (1994); Hall (1994); Edgington (1995) for overviews.

Figure 2: Distribution of values for *'if A then C'*

antecedent world can fall between 0 and 1, and it is not fixed: if the probability distribution changes, it will change along with it. For these reasons, the proposal has largely met with scepticism on conceptual grounds. Jeffrey and Stalnaker did not offer a compelling explanation of what these intermediate values are supposed to be values *of*.

Partly in an attempt to answer this question, Kaufmann (2005a) explored the possibility of interpreting the values at non-antecedent worlds as those of the corresponding counterfactual conditional. This perspective necessitates and amendment to the definition in (6) in view of the role of *causal dependencies* in the interpretation of counterfactuals. I will summarize the account in the next two subsections. The part of Jeffrey's intermediate-value approach that Kaufmann makes crucial use of is the distinction between the values of conditionals at individual worlds on the one hand, and their probabilities, on the other. The modification consists in the claim that the values at non-antecedent worlds should not in general coincide with the conditional probability, as they do by definition under the assignment in (6) above.

2.3 Counterfactuals

In search of a unified theory of all conditionals, many authors have commented on the connection between indicative conditionals and their counterfactual counterparts. Minimal pairs like those in (8) suggest that the difference may be no more than one in temporal reference: (8a) is unlikely now because, or to the extent to which, (8b) *was* unlikely at the time prior to the assassination at which its use would have been appropriate.[6]

(8) a. If Oswald had not killed Kennedy, someone else would have. *[now]*
b. If Oswald does not kill Kennedy, someone else will. *[11/21/63]*

[6] Notice also that this account is not at odds with the fact that the probabilities of (8a,b) may be quite different from that of (i) in light of the fact that Kennedy was in fact killed.

(i) If Oswald did not kill Kennedy, someone else did. *[now]*

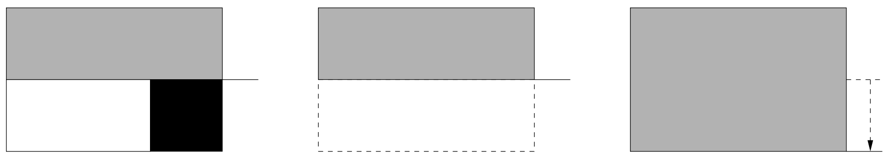

Figure 3: Update with the information that the number is odd

Likewise for the above (5), repeated once again as (9a), and its counterfactual counterpart (9b). In the event that the number is in fact odd, and the counterfactual becomes appropriate, its probability is intuitively the same as that of the indicative before the toss.[7]

(9) a. If she throws an even number, it will be a six.
  b. If she had thrown an even number, it would have been a six.

The value assignment defined above lends itself to the same interpretation. Figure 3 illustrates an update, through the aforementioned two-step procedure, with the information that the number is odd. At all remaining worlds, the value of the sentence is 1/3, the value of the indicative (9a) before the update.

The combination of the Thesis with the interpretation of counterfactuals as Past-tense forms of the corresponding indicatives was dubbed "Tense Probabilism" by Barker (1998). Barker went on to show that Tense Probabilism, while plausible in cases like the ones above, cannot be right in general. The argument relied on an example which had earlier been discussed by Slote (1978) and Bennett (1984), among others. To stay with the dice example discussed earlier, it can be summarized as follows.

Suppose Joe, a participant in the game, is asked before the toss (at Time 1) to place a bet on whether the outcome is even or odd. Joe believes correctly that the die is fair. In the scenario outlined in (10), his beliefs about the bet will be as indicated.

(10) a. Time 1: Before the bet is placed, since the die is fair,
  P(if Joe bets on 'odd', he will lose) $= .5$
  b. Time 2: Joe bets on 'even'.
  c. Time 3: The die is tossed and comes up even. At this point,
  P(if Joe had bet on 'odd', he would have lost) $= 1$

[7] I will not discuss the values of the counterfactual at worlds at which its antecedent is true. The assumption that it is equivalent to the material conditional in this case is shared by most conditional logics (Stalnaker, 1968; Lewis, 1973; Kratzer, 1981, and others).

The judgments in (10a,c) are hardly disputable. After the toss, the probability of the counterfactual is clearly different from that of the indicative at Time 1. In fact, there is *no* past time at which the probability of the indicative was 1. But this means, Barker concludes, that the probability of the counterfactual in (10c) cannot be identified with the prior probability of the indicative in (10a).[8]

2.4 Causality

Barker's argument is sound, but it refutes only one possible way of establishing a systematic connection between indicative and counterfactual conditionals. Recall that the account outlined above assigns two values to a conditional: its truth value and its probability. Kaufmann (2005a) suggests that while it is true that the sentences in (10a) and (10c) are not *equiprobable* at their respective times of evaluation, they are *equivalent*.

In view of judgments about counterfactuals, Kaufmann proposes to modify the value assignment by incorporating an account of causal independencies. The relevant fact about (10) above is that the outcome of the toss does not causally depend on the bet: At those worlds at which you bet on heads and win, the coin would still have come up heads even if you had bet on tails.

In general, such causal dependencies cannot be "read off" the probability distribution, but must be given separately as part of the model (cf. Woodward, 2001). This is easy to see if we consider a slight variant of the above scenario: Suppose there are two fair dice; die 1 is used in case Joe bets on 'even', die 2 if he bets on 'odd'. In this case, speakers generally agree that the probability of the counterfactual is as in (11c). However, the relevant probabilities are the same as before.

(11) a. Time 1: Before the bet is placed, since both dice are fair,
P(if Joe bets on 'odd', he will lose) = .5
b. Time 2: Joe bets on 'even'.

[8] Nor, Barker argues, are these judgments explained by the *prior propensity* account (Skyrms, 1981). This account differentiates between speakers' prior *subjective belief* in the indicative conditional and its *objective chance*. The relationship between the indicative and the counterfactual is then said to hold between their objective chances only, while prior subjective beliefs may be incorrect due to incomplete information.

The problem with this account is that in (10), Joe's posterior belief in the counterfactual would imply that he believes at time 3 that the prior objective probability of the indicative was 1, thus that the outcome was predetermined. But determinism is incompatible with Tense Probabilism: The posterior probability of (8a) above is not undefined, even though its antecedent is false.

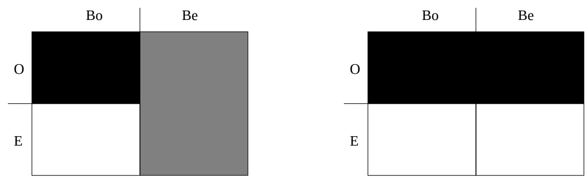

Figure 4: Value assignment for (10): global (left) and local (right). Bo/Be = bet on odd/even; O/E = the die comes up odd/even

c. Time 3: Die 1 is used and comes up even. At this point, P(if Joe had bet on 'odd', he would have lost) $= .5$

In Kaufmann's implementation, a *causal order* $\langle \Phi, \prec \rangle$ is a set of random variables (i.e., functions from possible worlds to numbers) ordered by the transitive and asymmetric relation $\prec$. The members of this set represent the "causally relevant" factors of the situation. The interpretation of the conditional depends on which factors are considered relevant and how they are related to each other. This is a source of ambiguity and context-dependence. I assume for simplicity that the members of Φ are equivalent to denotations of atomic sentences in the language, and futhermore, that they are few.[9]

For φ, ψ in Φ, the statement '$\varphi \prec \psi$' means that the expectation of ψ is determined by the value of φ (and possibly those of other variables). This causal order is used to modify the value assignment in (6): The value of *'if A then C'* at a world w at which A is false depends only on those A-worlds at which the variables that do not causally depend on $V(A)$ have the same values as they do at w. The definition is given in (12).[10]

(12) $$V_w(A \to C) = \begin{cases} V_w(C) \text{ if } V_w(A) = 1 \\ E[V(C)|V(A) = 1, \varphi = \varphi_w] \\ \quad \text{for all } \varphi \in \Phi \text{ such that } V(A) \not\preceq \varphi, \\ \quad \text{if } V_w(A) = 0 \end{cases}$$

In (10) above, the intuitively correct interpretation restults if we assume that the set of causally relevant variables includes the bet (Bo/Be for 'odd'

[9]This may not be reasonable if the goal is to give a metaphysically "true" representation of causal relations. It is plausible to assume, however, that speakers in their everyday use of conditionals consider no more than a few such factors, and ones that can be expressed in the language.

[10]Special provisions are required for the case that the falsehood of A is determined by the values of independent variables. I ignore this case here.

and 'even'), the outcome of the toss (O/E), and the winning or losing (W/L). As the scenario is set up, Bo/Be and O/E are causally independent of each other, and both jointly determine W/L. According to Definition (12), this affects the value of the conditional *'if Bo, then L'* at worlds at which its antecedent is false: The conditional expectation is only taken over those worlds at which the outcome of the toss is the same as at the world of evaluation. The effect of this restriction is an uneven distribution of values over the Bo-worlds, as shown on the right-hand side of Figure 4.

The role of causal relations in the interpretation of counterfactuals is increasingly being acknowledged in artificial intelligence, philosophy, and psychology (Pearl, 2000; Spirtes et al., 2000; Woodward, 2001, 2003; Glymour, 2001; Sloman and Lagnado, 2004). The interpretation given in (12) corresponds, in this framework, to the use of the 'do' operator introduced by Pearl.

2.5 Two probabilities

The dice example is of course uninteresting from a probabilistic point of view. The bet and the outcome of the toss jointly determine whether Joe wins or loses; there is no uncertainty once those facts are settled. I now turn to a more interesting scenario, a version of which was discussed by Kaufmann (2004), and which was also used in the experiment discussed below.

(13) You are about to pick a marble from a bag. There are two sorts of bags: X and Y.
   a. 75% are of type X: They contain ninety blue marbles and ten white ones.
   b. 25% are of type Y: They contain ten blue marbles and ninety white ones.

   In all bags, nine of the white marbles have a red spot.

Against the background of this scenario, consider the probability of (14).

(14) If the marble is white, it will have a red spot.

Nothing in the scenario suggests that the origin of the marble (i.e., whether the bag from which it is drawn is of type X or Y) depends on its color. This affects the value assignment according to Definition (12). Let 'X' be the statement that the marble is from bag X. The values of (14) are defined as in (15) (see Figure 5 for illustration).

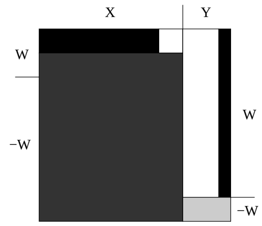

Figure 5: Assignment of values for (14)

$$
(15) \qquad V_w(\text{if } W \text{ then } S) = \begin{cases} V_w(S) \text{ if } V_w(W) = 1 \\ E[V(S)|V(W) = 1, V(X) = V_w(X)] \\ \qquad \text{if } V_w(W) = 0 \end{cases}
$$

$$
= \begin{cases} 1 \text{ if } V_w(W) = 1 \text{ and } V_w(S) = 1 \\ 0 \text{ if } V_w(W) = 1 \text{ and } V_w(S) = 0 \\ .9 \text{ if } V_w(W) = 0 \text{ and } V_w(X) = 1 \\ .1 \text{ if } V_w(W) = 0 \text{ and } V_w(X) = 0 \end{cases}
$$

Here as before, it is plausible to interpret the values assigned at those worlds at which the marble is not white as those of the corresponding counterfactual (16).

(16) If the marble had been white, it would have had a red spot.

Intuitively, this sentence is more likely if the bag is of type X than if it is of type Y. The values assigned according to (15) reflect this intuition.

2.6 Local and global interpretations

We are finally ready to see the fundamental tension in Kaufmann's account: While the assignment in (15) is sensible for the counterfactual in (16), the assumption that these same values are also those of the indicative in (14) is bluntly at odds with the central premise of the probabilistic account: Their expectation, given in (17), does not equal the conditional probability (18).[11]

---

[11] $E[V(\text{if } W \text{ then } S)]$
$= 1 \cdot P(SW) + 0 \cdot P(S\overline{W}) + P(S|WX) \cdot P(\overline{W}X) + P(S|W\overline{X}) \cdot P(\overline{W}\,\overline{X})$
$= P(SWX) + P(SW\overline{X}) + P(S|WX) \cdot P(\overline{W}X) + P(S|W\overline{X}) \cdot P(\overline{W}\,\overline{X})$
$= P(S|WX)[P(WX) + P(\overline{W}X)] + P(S|W\overline{X})[P(W\overline{X}) + P(\overline{W}\,\overline{X})]$

$P(S|W) = [P(SWX) + P(SW\overline{X})]/P(S)$
$= [P(W|SX)P(X|S)P(S) + P(W|S\overline{X})P(\overline{X}|S)P(S)]/P(S)$

(17) $$E[V(\text{if } W \text{ then } S)] = P(S|WX)P(X) + P(S|W\overline{X})P(\overline{X}) = .9 \cdot .75 + .1 \cdot .25 = \mathbf{.7}$$

(18) $$P(S|W) = P(S|WX)P(X|W) + P(S|W\overline{X})P(\overline{X}|W) = .9 \cdot .25 + .1 \cdot .75 = \mathbf{.3}$$

On the one hand, the Thesis maintains that the probabilities of indicative conditionals are the corresponding conditional probabilities, as in (18). On the other hand, the assumption about the relationship between indicative and counterfactual conditionals is that they are equivalent (though not equiprobable), and the above examples suggest that the proper value assignment for counterfactuals must be sensitive to causal relations. But then the probability of the indicative in the present scenario is predicted to be (17). On the face of it, it appears that the Thesis is incompatible with the proposed unified account of conditionals.

This conclusion is not inevitable, however. We can have it both ways if we can show that indicative conditionals have *both* of the probabilities in (17) and (18). This is in essence what Kaufmann (2004) claims: Rather than refuting the Thesis, he argues, such examples offer a deeper insight into a semantic variability of indicative conditionals between two interpretations, one "local" and the other "global." Kaufmann uses this hypothesis in the analysis of examples that have been proposed in the philosophical literature to point out discrepancies between the Thesis and intuitive probability judgments, and argues that moreover, under certain conditions it is *rational* to give a conditional its local interpretation. In addition, Kaufmann (2005b) argues that the account also yields superior probability assignments for embedded and compounded conditionals.

I will not review these discussions here in great detail. However, it is useful to clarify what exactly the difference corresponds to in terms of the intuition behind the Ramsey Test. Recall that the interpretation of conditionals involves two steps, elimination and renormalization. Kaufmann sees the difference between local and global interpetations in the way the second step is carried out.

Figure 6 illustrates. After the hypothetical update with the information that the marble is white, there are two ways of recalibrating the probabilities. Under the local interpretation, the relative probabilities of X and $\overline{X}$ are not affected. The probabilities in each of the cells in the X/$\overline{X}$-partition are calculated locally. In the resulting distribution (shown in the center), the

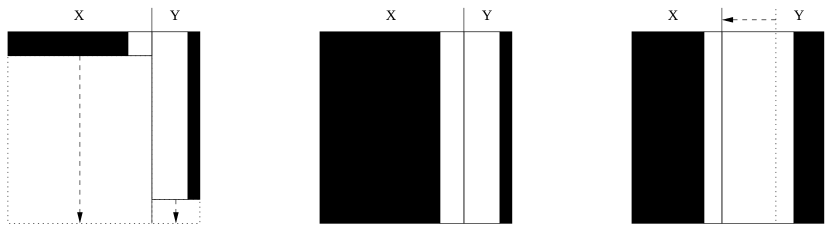

Figure 6: Renormalization after update with W (left): local (center) and global (right)

probability that the marble has a red spot (i.e., the black area in the figure) is large. Under the global interpretation, in contrast, the renormalization does affect X and $\overline{X}$: In the resulting probability distribution (shown on the right), their relative probabilities, too, have changed.

Formally, this difference corresponds to the use of $P(X)$ and $P(\overline{X})$ in (17), as opposed to $P(X|W)$ and $P(\overline{X}|W)$ in (18), as weights. (Notice that the *values* of the conditional are the same under both interpretations: Only the weights change.) Intuitively, the inference involved in the local interpretation can be paraphrased as follows:

(19) a. In the X bags, most of the white marbles have a red spot.
b. In the Y bags, few of the white marbles have a red spot.
c. There are more X bags than Y bags.
d. The probability of the conditional is more likely high than low.

The global interpretation, on the other hand, corresponds to the inference in (20). The crucial difference lies in the *abductive* step highlighted in (20c).

(20) a. Suppose the marble is white.
b. There are many white marbles in Y bags and few in X bags.
c. *So it is probably from a Y bag.*
d. In the Y bags, few of the white marbles have a spot.
e. Then the marble probably won't have a spot.
f. So the probability of the conditional is low.

Both of (19) and (20) appear sensible, even though they lead to opposite conclusions.

3 Experiments

The above account is in line with psychological evidence that causal relations are important in interpreting not only counterfactuals, but also indicative conditionals (Sloman and Lagnado, 2001, 2004; Sloman, p.c.). In this section I describe a preliminary pilot study (partly reported in Kaufmann et al., 2004) to illustrate an experimental paradigm within which I will test its predictions as part of this project.

An observable difference between local and global interpretations is predicted whenever the probability of the consequent depends, in addition to the antecedent, on a variable that is causally independent but stochastically dependent upon the latter. In the above example, this third variable is the type of the bag (X or Y), and the local and global probabilities are 0.7 and 0.3, respectively.

One hypothesis one may base on this is that subjects' judgments, when presented with a scenario like the one above and asked to assess the probability of the conditional, will exhibit a bimodal distribution, corresponding to the two readings. However, such an observation alone would not constitute conclusive evidence for the account. The scenario is complicated, and variation across subjects may simply be due to confusion. A better question to ask is whether responses across stimuli vary systematically with the manipulation of relevant variables. According to the theoretical account, which interpretation is chosen depends on two factors: (i) whether subjects perform the inference step that is absent in the local interpretation (cf. 19) but present in the global one (cf. 20); and (ii) whether the third variable in the scenario is causally dependent on the antecedent or not.

3.1 Experiment 1.

The primary goal of the first experiment was to test whether subjects' judgments are affected when the abductive inference step is primed by the context in which the target conditional is presented. In addition, two posterior probabilities were elicited: that of the consequent upon learning that the antecedent is true, and that of the counterfactual upon learning that the antecedent is false. In sum, the following predictions were tested:

C1: Subjects' probability judgments will show a bimodal distribution, split between *'likely'* and *'unlikely'* for the local and global interpretations, respectively. If one of these interpretations is preferred, this bias will be reflected in the distribution of responses.

C2: We expect a higher incidence of global (*'unlikely'*) responses, compared to C1, if the sentence is presented in a context in which the abductive inference step is made salient, as in (21).

(21) a. If the marble is white, it will be from a Y-bag.
b. If the marble is white, it will have a red spot.

P: Under the Ramsey Test, the update with the antecedent is *hypothetical*. It is often assumed that the *permanent* update upon learning that the antecedent is true proceeds by conditionalization (Lewis, 1976). If this is true, judgments for the consequent (22) in a context in which subjects have been told that the marble has been drawn and is white, will correspond to the global interpretation.

(22) The marble will have a red spot.

Cf: Counterfactuals provide an important part of the motivation for the theory. We expect judgments about the counterfactual (23), in a context in which subjects have been told that the marble has been drawn and is not white, to conform to the local interpretation (*'likely'*), since in this case the bag is more likely to be of type X, where 90% of the white marbles have a red spot.

(23) If the marble had been white, it would have had a red spot.

3.1.1 Method

55 undergraduate students of Northwestern University participated in the study as part of a course requirement. Subjects were given a questionnaire in which the scenario was described on four pages, one for each condition. Subjects were instructed not to refer back to previous responses as they moved on. The descriptions of the scenario were almost idential on each page, except for the statement that a certain had been drawn in Conditions P and and Cf. On each page, the scenario was followed by three sentences including the target conditional and two filler items (however, in C2, the sentence preceding the target was the conditional used to prime the abductive inference). Subjects were asked to assess the probability of each sentence based on the information provided, by circling an item on the scale *'likely'*, *'fifty-fifty'*, *'unlikely'*, *'don't know'*. They were instructed to consult only

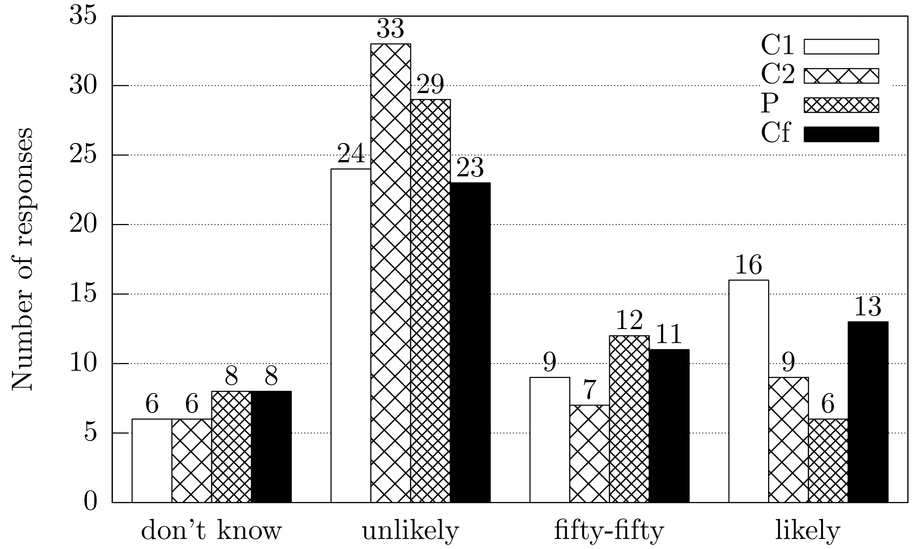

Figure 7: Responses in Experiment 1

their linguistic intuitions, not using the numbers given in the scenario for calculations.

3.1.2 Results

The results are summarized in Figure 7. With regard to the predictions mentioned earlier, the following was found:

C1: Responses show a bimodal distribution, with peaks at '*unlikely*' and '*likely*'. The higher incidence of '*unlikely*' responses suggests a bias towards the global interpretation.

C2: There was a significant tendency for subjects to judge the conditional less likely in C2 than in C1 (Wilcoxon:[12] $P = 0.0226$).

P: Judgments of the probability of the consequent upon learning that the antecedent is true differed significantly from C1 ($P = 0.0328$) but not from C2 ($P = 0.9461$).

[12] The Wilcoxon matched-pairs signed-ranks test compares judgments pairwise within subjects to determine whether there is a general upward or downward trend.

Cf: The judgments for the counterfactual did not differ significantly from those for C1 ($P = 0.5975$), as predicted. The prediction that they would differ from C2 is only weakly supported: The difference is greater than that to C1, but not significant ($P = 0.1333$).

3.1.3 Discussion

The bimodal distribution observed in Condition C1 is compatible with the hypothesis that conditionals have both interpretations, although as mentioned above, it does not constitute conclusive evidence. With this caveat, the preponderance of '*unlikely*' responses suggests that the global interpretation is the preferred one. The difference between C1 and C2 suggests that the global interpretation is preferred when the target follows a sentence which explicitly invokes the inference from the color of the marble to the type of the bag. Furthermore, the patterns in Conditions P and Cf lend some support to the predictions about posterior probabilities outlined above.

That said, the materials used in this experiment had some conspicuous weaknesses. Most importantly, the order in which the four conditions appeared was constant. This is especially problematic for C1 and C2: The higher incidence of '*unlikely*' responses (suggesting global interpretations) may have been merely an ordering effect, since C2 always followed C1. The results of Experiment 2 below, in which the order was counterbalanced, do not fully resolve that issue. Furthermore, the marble scenario is abstract and somewhat artificial. While this has the advantage of avoiding the interference of subjects' world knowledge, some subjects reported in debriefing that they had tried to calculate based on the numbers, in spite of the instructions not to do so. In a separate experiment (not reported here for lack of space), this problem was addressed by showing subjects real bags with actual marbles of various colors. The results obtained in this way were similar to the ones presented here. Finally, some subjects found the scale on which they marked their judgments problematic: It did not include an option for certain truth or falsehood, and for some the label '*fifty-fifty*' meant "uncertain" in a very general sense, instead of "probability of ca. 0.5" as intended.

3.2 Experiment 2

A second experiment tested whether subjects' responses depend on the causal structure in the scenario. The same distribution of marbles was described in two different ways: The "static" description was the same as for Experiment 1 (following (13) above). The alternative was a "dynamic" description suggesting that the bags were filled according to a rule, whereby marbles were placed in different bags according to their colors:

(24) One hundred marbles were filled in each bag.
  a. 70% of the marbles are black.
     Ninety black marbles were put in each X-bag;
     Ten black marbles were put in each Y-bag.
  b. 30% of the marbles are white.
     Ten white marbles were put in each X-bag, nine of them with a red spot;
     Ninety white marbles were put in each Y-bag, nine of them with a red spot.

Recall that the theory predicts a difference between local and global interpretations only when there is no such dependency. Thus we expect that subjects tend to interpret the conditional globally in the dynamic condition, resulting in a higher incidence of *'unlikely'* responses.

3.2.1 Method

25 undergraduate students enrolled in Northwestern's summer program participated for a small financial compensation. They were given a questionnaire similar to the one in Experiment 1. Only Conditions C1 and C2 were tested. Each questionnaire contained two blocks, one Dynamic and one Static, separated by twelve pages of unrelated filler scenarios. The order of the Dynamic and Static blocks was counterbalanced; likewise, within each block, the order of the conditions C1 and C2 was counterbalanced. However, where C1 preceded C2 in the first block, C2 preceded C1 in the second block and *vice versa*, thus there were a total of four different versions. The scale on which subjects marked their judgments was *'can't say'*, *'false'*, *'unlikely'*, '' (empty), *'likely'*, and *'true'*. Otherwise, the instructions were similar to those in Experiment 1.

3.2.2 Results

The results are shown in Figure 8. The Static condition (left) shows a bimodal distribution for C1, with a bias towards the *'likely'* (local) response. In C2, the bias is reversed, although this does not involve a higher frequency of *'unlikely'* responses. The Dynamic condition shows a bias for *'unlikely'*

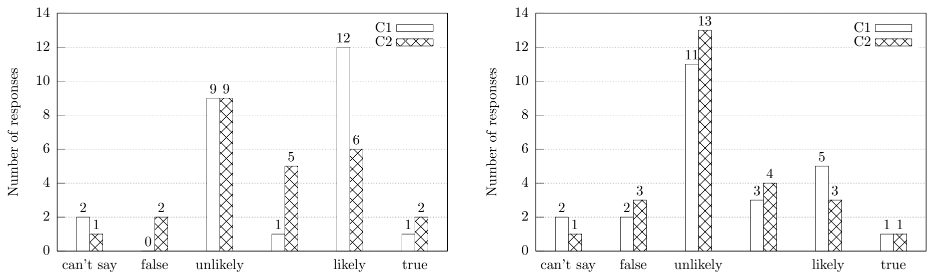

Figure 8: Responses in Experiment 2, Static (left) and Dynamic (right) conditions

responses under both conditions, slightly stronger in C2 than in C1. However, none of these trends was significant, given the small number of subjects (N=25; cf. 55 in Experiment 1).

3.2.3 Discussion

The bias towards *'unlikely'* responses is in line with the hypothesis that the dynamic condition, unlike the static one, does not induce a local interpretation for the conditional. The results show a trend away from *'likely'* (local) responses in Condition C2, where the target conditional follows a sentence which primes the abductive inference. This was also observed in Experiment 1; however, in the static condition, it did not translate into a higher incidence of *'unlikely'* judgments. The interpretation of this fact is difficult, given the small number of subjects. Finally, the above comments on the idiosyncrasies of the marble scenario apply likewise to this experiment.

4 Overall discussion and future directions

Further work is needed to corroborate these preliminary results. The studies do suggest, however, that the experimental paradigm is suitable for testing the predictions of the theory. Much of the work in the initial stages of the project will be aimed at refining the experiment design and the development of a larger and more diverse set of stimuli. In this, I will benefit from consultations and collaborations with colleagues, especially Lance Rips (Psychology, Northwestern), the author of influential psychological studies on conditionals and reasoning (Rips and Marcus, 1977; Rips, 1996), from whose comments this work has benefited and with whom I will be co-teaching a seminar on this topic in Year 1 (see the attached letter). Based on this preparatory work and with support from the grant, I will conduct a series of related experiments on a larger scale in Years 1 through 3.

These studies also raise a more "linguistic" question: How does the way in which conditionals are interpreted depend on the semantic properties of the sentences involved? Recall that the difference between the static and dynamic conditions in Experiment 2 lay in the language in which the scenario was described. Since these descriptions were the only source of information available to subjects, any differences in their responses must at bottom be due to semantic differences, truth-conditional or otherwise, between the sentences used in the description. What are the relevant linguistic differences? In the latter stages of the project, I will explore the hypothesis that subjects' preferences correlate with the aspectual and thematic properties used in the description. Dowty (1979) proposed that the difference between certain aspectual classes is attributable in part to the presence or absence of an abstract operator 'CAUSE', whose model-theoretic interpretation is inspired by the counterfactual theory of causality of Lewis (1979) and thus indirectly related to my analysis of conditionals. Other lexical distinctions which have been found to involve the CAUSE operator include unergatives *vs.* unaccusatives (Pustejovsky, 1991) and agent *vs.* patient proto-roles in argument structure (Dowty, 1991). In the present context, this suggests a connection between lexical semantics and subjects' judgments. Specifically, the relevant causal relations may be present (giving rise to local interpretations for indicatives and driving certain counterfactual inferences) just when the sentences used in the description involve the CAUSE operator, and absent otherwise. These are very tentative ideas at this point; but the work discussed here has established a theoretical framework in which to formulate such predictions and an experimental method for testing them.

Bibliography

Barker, S. J. 1998. Predetermination and tense probabilism. *Analysis*, 58: 290–296.

Bennett, J. 1984. Counterfactuals and temporal direction. *The Philosophical Review*, 93(1):57–91.

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

Dowty, D. 1979. *Word Meaning and Montague Grammar*, volume 7 of *Studies in Linguistics and Philosophy*. Reidel.

Dowty, D. 1991. Thematic proto-roles and argument selection. *Language*, 67:547–619.

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

Eells, E. and B. Skyrms, editors. 1994. *Probabilities and Conditionals: Belief Revision and Rational Decision*. Cambridge University Press.

Galavotti, M. C., P. Suppes, and D. Costantini, editors. 2001. *Stochastic Causality*. CSLI Publications.

Gärdenfors, P. 1982. Imaging and conditionalization. *Journal of Philosophy*, 79:747–760.

Gibbard, A. 1981. Two recent theories of conditionals. In Harper et al. (1981), pages 211–247.

Glymour, C. 2001. *The Mind's Arrows: Bayes Nets and Graphical Causal Models in Psychology*. The MIT Press.

Hájek, A. and N. Hall. 1994. The hypothesis of the conditional construal of conditional probability. In Eells and Skyrms (1994), pages 75–110.

Hall, N. 1994. Back in the CCCP. In Eells and Skyrms (1994), pages 141–160.

Harper, W. L., R. Stalnaker, and G. Pearce, editors. 1981. *Ifs: Conditionals, Belief, Decision, Chance, and Time*. Reidel.

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. 2004. Conditioning against the grain: Abduction and indicative conditionals. *Journal of Philosophical Logic*, 33(6):583–606.

Kaufmann, S. 2005a. Conditional predictions: A probabilistic account. *Linguistics and Philosophy*, 28(2):181–231.

Kaufmann, S. 2005b. Conditionals right and left: Probabilities for the whole family. Ms., Northwestern University. http://ling.northwestern.edu/~kaufmann/Papers/rightleft.pdf.

Kaufmann, S., E. Winston, and D. Zutty. 2004. Local and global interpretations of conditionals. In *Proceedings of the Eighth Symposium on Logic and Language (LoLa8)*, Debrecen, Hungary. To appear.

Kratzer, A. 1981. Partition and revision: The semantics of counterfactuals. *Journal of Philosophical Logic*, 10:201–216.

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

Lewis, D. 1976. Probabilities of conditionals and conditional probabilities. *Philosophical Review*, 85:297–315.

Lewis, D. 1979. Counterfactual dependence and time's arrow. *Noûs*, 13: 455–476.

Lewis, D. 1986. Probabilities of conditionals and conditional probabilities II. *Philosophical Review*, 95:581–589.

McGee, V. 2000. To tell the truth about conditionals. *Analysis*, 60:107–111.

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

Oaksford, M. and N. Chater. 2003. Conditional probability and the cognitive science of conditional reasoning. *Mind & Language*, 18(4):359–379.

Over, D. E. and J. S. B. T. Evans. 2003. The probability of conditionals: The psychological evidence. *Mind & Language*, 18(4):340–358.

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

Pollock, J. L. 1976. The "possible worlds" analysis of counterfactuals. *Philosophical Studies*, 29:469–476.

Pustejovsky, J. 1991. The syntax of event structure. *Cognition*, 41:47–81.

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

Rips, L. J. 1996. *The Psychology of Proof*. MIT Press, Cambridge, MA.

Rips, L. J. and S. L. Marcus. 1977. Suppositions and the analysis of conditional sentences. In Just, M. A. and P. A. Carpenter, editors, *Cognitive Processes in Comprehension*, pages 185–220. Lawrence Erlbaum, Hillsdale, NJ.

Skyrms, B. 1981. The prior propensity account. In Harper et al. (1981), pages 259–265.

Skyrms, B. 1984. *Pragmatics and Empiricism*. Yale University Press.

Skyrms, B. 1994. Adams conditionals. In Eells and Skyrms (1994), pages 13–26.

Sloman, S. A. and D. A. Lagnado. 2001. Counterfactual undoing in determinstic causal reasoning. In *Proceedings of the 24th Annual Conference of the Cognitive Society*.

Sloman, S. A. and D. A. Lagnado. 2004. Do we 'do'? To appear in *Cognitive Science*.

Slote, M. A. 1978. Time in counterfactuals. *The Philosophical Review*, 87 (1):3–27.

Spirtes, P., C. Glymour, and R. Scheines. 2000. *Causation, Prediction, and Search*. MIT Press.

Stalnaker, R. 1968. A theory of conditionals. In *Studies in Logical Theory, American Philosophical Quarterly, Monograph: 2*, pages 98–112. Blackwell.

Stalnaker, R. 1975. Indicative conditionals. *Philosophia*, 5:269–286.

Stalnaker, R. and R. Jeffrey. 1994. Conditionals as random variables. In

Eells and Skyrms (1994), pages 31–46.
Woodward, J. 2001. Probabilistic causality, direct causes and counterfactual dependence. In Galavotti et al. (2001), pages 39–63.
Woodward, J. 2003. *Making Things Happen: A Theory of Causal Explanation.* Oxford University Press.