LOCAL AND GLOBAL INTERPRETATIONS OF CONDITIONALS

STEFAN KAUFMANN
ERICA WINSTON
DEBORAH ZUTTY

Department of Linguistics
Northwestern University
{kaufmann, e-winston, d-zutty}@northwestern.edu

The unified probabilistic theory of indicative and counterfactual conditionals proposed by Kaufmann (2005) 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 probability of a conditional *'if A, C'* is the conditional probability of C, given A, has a long history. It was hinted at in Ramsey's (1929) suggestion that conditionals are used to argue about "degrees of belief," and philosophers have since extensively studied its ramifications for logical and semantic theory. We will not discuss these developments here (see Eells and Skyrms, 1994 and Edgington, 1995 for recent overviews). Our goal is rather to present experimental evidence in support of Kaufmann's (2004) claim that while the idea is basically right, it must be modified if it is to account for a wider range of empirical facts.

We will begin with an outline of the relevant theoretical considerations (Sections 2 and 3) before describing our experiment (Section 4). Due to constraints on space, this preliminary discussion will be brief. The reader is referred to the works cited for further details and motivation.

2 Background

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

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

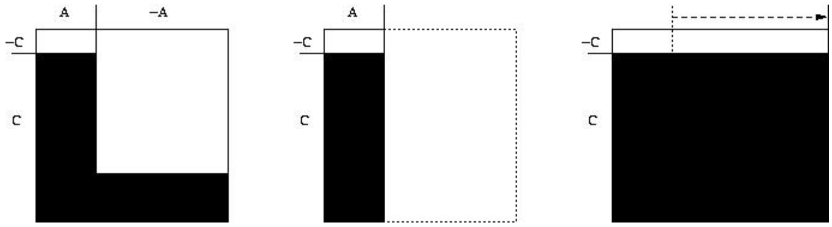


Figure 1: Update with A in two steps: elimination and renormalization

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 given explicit formal representations. The probabilistic framework we adopt makes the following assumptions:

(1) 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.

We 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 update of Pr with the information that A is true plays an important part below and deserves some discussion. It is useful to think of this operation as involving two steps: *elimination* and *renormalization*. The former eliminates those possible worlds at which A is false. The latter consists in recalibrating the probabilities of the remaining sets of worlds (at which A is true) in such a way that the relative probabilities of all sentences other than A are preserved.

This procedure may be visualized as in Figure 1. After the elimination of those possibilities at which A is false, the probabilities are recalculated in such a way that A receives probability 1 and the proportions within the set of A-worlds are restored. The third picture shows the result. The probability of the C-worlds, shown as the striped area in the picture, is larger after the update than it was in the original distribution. This is the *posterior* probability of C. The assumption underlying the probabilistic approach is that it equals the *prior* probability of the conditional *'if A, then C'*.

Pr is a probability distribution over propositions, but we are interested in the sentences which denote those propositions. To make this connection, we follow Kaufmann (2005). Atomic sentences and truth-functional compounds thereof are assigned truth values at possible worlds by a valuation function V. Based on V and Pr, we define a

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

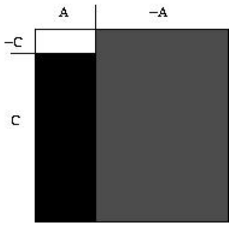


Figure 2: Distribution of values for *'if A then C'*

probability distribution P over such sentences. The probability $P(\varphi)$ of a sentence φ is the *expectation* of its truth value, written '$E[V(\varphi)]$' and defined as the weighted sum of its values, where the weights are the probabilities that $V(\varphi)$ has those values:[3]

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

The extension of V to conditionals is not entirely straightforward. Lewis (1976) showed that conditional probabilities are not probabilities of propositions. We cannot in general define an assignment $V(\text{if } A, \text{then } C)$ in such a way that the expectation of its values is guaranteed to equal the conditional probability of C, given A, for any probability distribution Pr. We *can*, however, define a value assignment which depends on Pr:

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

Thus at worlds at which A is true, the conditional is equivalent to the material conditional. At worlds at which A is false, the value of the conditional is the conditional expectation of C, given that A is true. This value equals the conditional probability of C, given A, and may fall anywhere in the interval $[0, 1]$. The resulting distribution of values for the conditional is illustrated in Figure 2 (the colors correspond to values: black = 1; white = 0; grey = intermediate).

2.2 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 (4) suggest that the difference may be no more than one in temporal reference: (4a) is unlikely now because, or to the extent to which, (4b) *was* unlikely at the time prior to the assassination at which its use would have been appropriate.

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

[3] We 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.

The value assignment defined above and schematically illustrated in Figure 2 lends itself to an interpretation that accords well with this view. Figure 2 shows the distribution of values before the possibility that the antecedent be true is eliminated.[4] Suppose this distribution is updated, through the aforementioned two-step procedure, with the information that A is false. Then at all remaining worlds, the value of *'if A then C'* is the prior conditional probability of C, given A, hence the value that the indicative *had* before the update.

The combination of the probabilistic approach with the thesis that counterfactuals are Past-tense forms of the corresponding indicatives was labeled "Tense Probabilism" by Barker (1998). Barker argues that Tense Probabilism, while plausible in cases like (4), cannot be right in general. To show this, he uses the following example, which had earlier been discussed by Slote (1978) and Bennett (1984), among others.

(5) a. Time 1: An unbiased coin is about to be tossed. You are asked to make a bet. At this point, since the coin is fair,
$P(\text{if I bet on tails, I will lose}) = .5$
b. Time 2: You bet on heads.
c. Time 3: The coin is tossed and comes up heads. At this point,
$P(\text{if I had bet on tails, I would have lost}) = 1$

The judgments in (5a,c) are hardly disputable. But this means, Barker argues, that the counterfactual in (5c) cannot be considered the Past-tense version of the indicative in (5a).[5]

2.3 Causality

Barker's argument against Tense Probabilism refutes only one possible way of establishing a systematic connection between indicative and counterfactual conditionals. Kaufmann (2005) argues that while it is true that the sentences in (5) have different probabilities at their respective times of evaluation, this does not imply that they are not equivalent. The solution Kaufmann proposes makes crucial use of an assignment of values to conditionals that is sensitive to *causal independencies*. This notion will also be central to the data we discuss below.

The basic idea, with regard to (5) above, is to incorporate the intuition 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. Instead, they must given as part of the model. Specifically, 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 are intended to 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

[4] What exactly this means depends on the interpretation of the probability distribution Pr. It may be the subjective probability entertained by a speaker who is uncertain as to the true of A; alternatively, it may represent the objective chances of future events at a time at which both A and $\overline{A}$ are open possibilities.

[5] Nor, Barker argues, are these judgments explained by the *prior propensity* account (Skyrms, 1981).

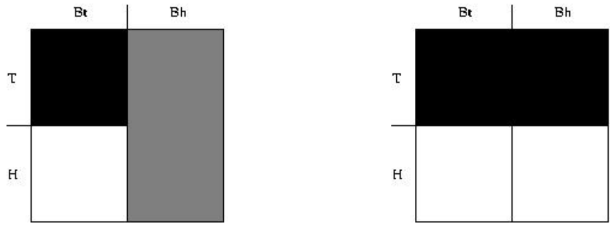


Figure 3: Value assignment for (5): global (left) and local (right). Bt/Bh = you bet on tails/heads; T/H = the coin comes up tails/heads

other. This is a source of ambiguity and context-dependence. We will assume here for simplicity that the members of Φ are denotations of atomic sentences in the language, and futhermore, that they are few.[6]

For φ, ψ in Φ, the statement that $\varphi \prec \psi$ means that the expectation of ψ is determined by the value of φ (and possibly those of other variables). This relation is used to modify the value assignment in (3): 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 (6).[7]

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

In (5) above, the intuitively correct interpretation follows if we assume that the set of causally relevant variables includes the bet (Bh/Bt), the outcome of the toss (H/T), and the winning or losing (W/L). As the scenario is set up, Bh/Bt and H/T are causally independent of each other, and both jointly determine W/L. According to Definition (6), this affects the value of the conditional *'if Bt, 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 Bh-worlds, as shown on the right-hand side of Figure 3.

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; Galavotti et al., 2001; Sloman and Lagnado, 2004; Woodward, 2003). The interpretation given in (6) corresponds, in our framework, to the use of the 'do' operator introduced by Pearl.

[6]This may not be reasonable if the goal is to give a metaphysically "true" representation of causal relations. We do think it plausible, however, the speakers in their everyday use of conditionals consider no more than a few such factors.

[7]Special provisions are required for the case that the falsehood of A is *entailed* by the values of independent variables. We ignore this case here.

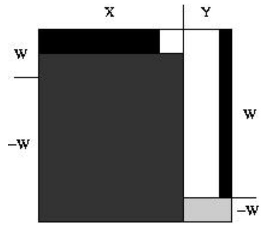


Figure 4: Assignment of values for (8)

3 Local and global interpretations

The coin example is of course uninteresting from a probabilistic point of view. The bet and the outcome of the toss jointly determine whether you win or lose; there is no uncertainty once those values are known. We now turn to a more interesting scenario, which we also used in our experiment.

(7) 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, we are interested in the probability of (8).

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

We take it to be clear that in this scenario the origin of the marble (i.e., whether the bag from which it is drawn is of type X or Y) does not depend on its color. This affects the value assignment according to Definition (6). Let 'B' denote the variable "bag," whose value can be either X or Y. The values of (8) are defined as in (9) (see Figure 4).

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

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 (10).

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

(10) is more likely if the bag is of type X, than if it is of type Y. The values assigned according to (9) are .9 and .1, respectively, which seems intuitively right in this scenario.

3.1 The problem

Now it turns out that the assumption that these values are also assigned to the indicative conditional (8) is at odds with the central premise of the probabilistic account: The expectation of those values (11) does not equal the conditional probability (12).[8]

(11) $$\begin{aligned} E[V(\text{if } W \text{ then } S)] =& P(S|WX)P(X) + P(S|WY)P(Y) \\ =& .9 \cdot .75 + .1 \cdot .25 = \mathbf{.7} \end{aligned}$$

(12) $$\begin{aligned} P(S|W) =& P(S|WX)P(X|W) + P(S|WY)P(Y|W) \\ =& .9 \cdot .25 + .1 \cdot .75 = \mathbf{.3} \end{aligned}$$

This is the problem our experiment addresses. To recapitulate: On the one hand, the standard probabilistic account maintains that the probabilities of indicative conditionals are the corresponding conditional probabilities, as in (12). On the other hand, our assumption about the relationship between indicative and counterfactual conditionals is that they are equivalent (though not equiprobable), and the 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 (11). Has the attempt to give a unified account of conditionals reached a dead end?

3.2 The hypothesis

The problem is not as serious as it seems — in fact, there is no problem. Kaufmann (2004) discusses discrepancies like the one between (11) and (12) and argues that they actually offer a deeper insight, namely that indicative conditionals are *ambiguous* between two readings, one "local" and the other "global." Kaufmann shows that this hypothesis explains a number of cases that have been proposed in the philosophical literature as counterexamples to the claim that the probabilities of conditionals are the corresponding conditional probabilities, and furthermore, that under certain conditoins it is *rational* to give a conditional its local interpretation.

We will not review those arguments here. 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 5 shows what is meant by this. After the temporary 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 Y are not affected. The probabilities in each of the cells in the X/Y-partition are calculated locally. In the resulting distribution (shown in the center), the 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 Y: In the resulting probability distribution (shown on the right), their relative probabilities, too, have changed.

---

[8] $$\begin{aligned} E[V(\text{if } W \text{ then } S)] &= P(SW) + P(S|WX)P(\overline{W}X) + P(S|WY)P(\overline{W}Y) \\ &= P(S|WX)[P(WX) + P(\overline{W}X)] + P(S|WY)[P(WY) + P(\overline{W}Y)] \\ P(S|W) &= [P(SWX) + P(SWY)]/P(S) \\ &= [P(W|SX)P(X|S)P(S) + P(W|SY)P(Y|S)P(S)]/P(S) \end{aligned}$$

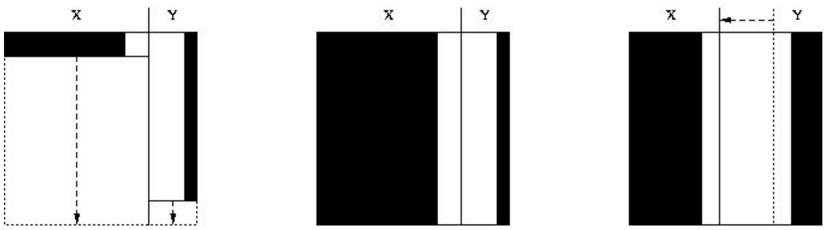


Figure 5: Renormalization after update with W (left): local (center) and global (right)

Formally, this difference corresponds to the use of $P(X)$ and $P(Y)$ in (11), as opposed to $P(X|W)$ and $P(Y|W)$ in (12), as weights. (Notice incidentally 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:

(13) 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 (14). The crucial difference lies in the *abductive* step (to the most likely explanation) highlighted in (14c).

(14) 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.

To sum up, the apparently contradictory predictions of Kaufmann's probabilistic theory of conditionals have upon closer inspection found a plausible explanation, one which crucially relies on a rarely noticed variability in the interpretation of indicatives. But does this variability really exist? Or have we bent our own intuitions to suit our theory-internal concerns? This is the question we will address in the remainder of this paper.

4 Experiment

The purpose of the experiment[9] is to test four predictions of the hypothesis outlined above:

Prediction 1 (C1). If conditionals are ambiguous in the way described above, subjects' probability judgments will show a bimodal distribution. In the marble scenario above,

[9]Due to constraints on space, we report here only the results of the first in a series of experiments. Further results, which basically corroborate these findings, will be discussed in an expanded version of the paper.

where the local and global interpretations of (8) yield appreciably different probabilities (.7 and .3, respectively), judgments will vary between '*likely*' and '*unlikely*'. In addition, if one of the interpretations is preferred, this bias will be reflected in the distribution of responses.

Prediction 2 (C2). What distinguishes local and global interpretations is the abductive inference step highlighted in (14) above. We therefore expect an increased number of global ('*unlikely*') responses, compared to Condition 1, if the sentence is placed in a context in which this abductive step is made salient, as in (15).

(15) 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.

Prediction 3 (P). The update with the antecedent is *hypothetical*. We do not make any claims as to how the *permanent* update, upon learning that the antecedent is true, is carried out. It is often assumed, however, that the latter proceeds by conditionalization (Lewis, 1976). If this is the case, the judgments for the consequent (16) in a context in which the subjects have been told that the marble has been drawn and is white, will correspond to the global interpretation.

(16) The marble will have a red spot.

Prediction 4 (Cf). As discussed in Section 2, counterfactuals provide an important part of the motivation of our account. We expect judgments about the counterfactual in (17), 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.

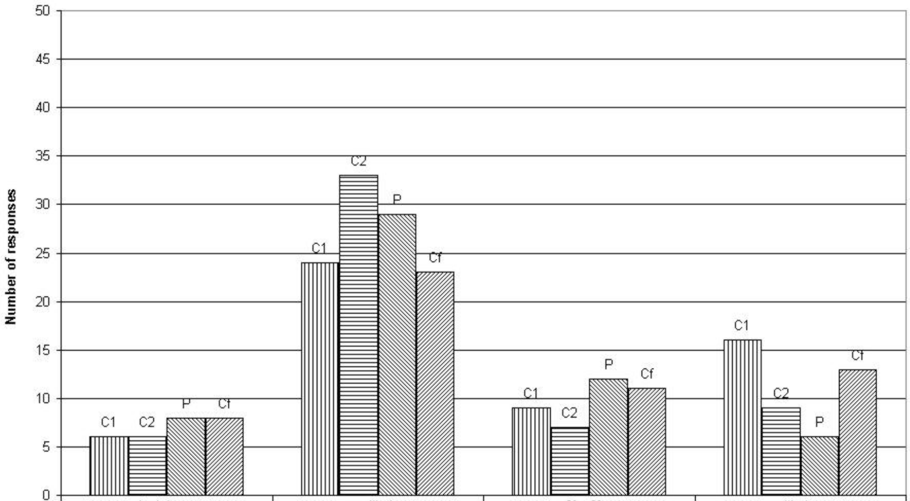
(17) If the marble had been white, it would have had a red spot.

4.1 Method

Fifty-five undergraduate students of Northwestern University participated in the study for course credit. Subjects were given a questionnaire in which the scenario was described on four pages, one for each condition. They were instructed not to refer back to previous responses as they moved on. The descriptions of the scenario were almost identical, except for the statement that a certain marble has been drawn in Conditions P and Cf. In each condition, the description of the scenario was followed by three sentences containing the target conditional and two filler items (however, in Condition C2, the sentence preceding the target was the conditional used to prime the abductive inference). Subjects were then asked to assess the probability of each sentence based of the information provided, by circling an item on the scale '*likely*, *fifty-fifty*, *unlikely*, *don't know*'. All subjects were asked to consult only their linguistic intuitions, rather than using the numbers given in the scenario for calculations.

4.2 Results

The data for conditions 1–4 are summarized in Figure 6.



	don't know	unlikely	fifty-fifty	likely
C1	6	24	9	16
C2	6	33	7	9
P	8	29	12	6
Cf	8	23	11	13

Figure 6: Results for conditions 1–4

C1. The bimodal distribution for this condition clearly supports our claim that the conditional is ambiguous. In addition, we observe a higher incidence of *'unlikely'* judgments, indicating a bias towards the global interpretation.

C2. We found a significant tendency for subjects to judge the conditional less likely in C2 than in C1 (Wilcoxon:[10] $P = 0.0226$).

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 ($P = 0.9461$).

Cf. As shown in Figure 6, the judgments for counterfactuals do not differ significantly from those for C1 ($P = 0.5975$). Our prediction that they would differ from C2 is supported, but only weakly: While this difference is greater than that to C1, it is not statistically significant ($P = 0.1333$).

4.3 Discussion

Our findings confirm the hypothesis that conditionals are ambiguous in the way described above. The global interpretation (under which the probability of the conditional is the corresponding conditional probability) turns out to be the preferred one. Our assumption

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

that the global interpretation differs from the local one in the presence of an abductive inference step is supported by the observation that the former can be primed in a context in which this abductive step is made explicit.

These results represent only one of a series of experiments using a more varied set of scenarios and stimuli. We do not present the full range of the data in this version of the paper, but they do confirm the conclusions we have reached here. They also address a couple of rather obvious shortcomings of the present setup, which we would briefly like to comment on.

First, the order in which the four conditions were presented did not differ between subjects. This is problematic especially for C1 and C2, where we observed a higher incidence of global interpretations in the latter. Based on the data we discuss here, the possibility cannot be ruled out that this is merely an ordering effect, since C2 always follows C1. However, in a later version of the questionnaire, we counterbalanced the two conditions and obtained similar overall results.

Second, the marble scenario is abstract and somewhat artificial in its use of numbers. This has the advantage of avoiding the interference of subjects' world knowledge. On the other hand, some subjects reported after the experiment that they had tried to calculate conditional probabilities based on the numbers, in spite of the instruction not to do so. In another experiment, we addressed this problem by showing subjects actual bags with marbles of various colors. The results we obtained in this way were similar to the ones presented here.

5 Conclusion

Although the range of data we discussed is limited, the results provide good empirical evidence for the theory of conditionals we outlined in Sections 2 and 3 above. To recapitulate once again, we started out with the common probabilistic interpretation of the Ramsey Test, according to which their probabilities are the corresponding conditional probabilities. On the other hand, the treatment of counterfactuals in terms of causal independencies implies that the prior expectation of their truth values cannot always equal the conditional probability. Finally, the third claim is that indicatives and their counterfactual counterparts are equivalent. We showed that if all three of these assumptions are to be true, then indicative conditionals must have (at least) two possible readings.

The results suggest that conditionals do indeed have these readings. If the account holds up to further scrutiny and exploration, it will provide an important missing piece in the development of a the unified theory of conditionals.

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.

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.

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

Kaufmann, S. 2004. Conditioning against the grain: Abduction and indicative conditionals. To appear in *Journal of Philosophical Logic*.

Kaufmann, S. 2005. Conditional predictions: A probabilistic account. To appear in *Linguistics and Philosophy*.

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

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

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

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

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

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.

Woodward, J. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press.