Probabilistic language in indicative and counterfactual conditionals

Daniel Lassiter, Stanford University

Abstract of poster to be presented at Semantics & Linguistic Theory 2017, University of Maryland

In the scene below, each of (1-3) is true against the specified informational background.

(1) [No relevant information] If the ball is from B, it’s likely to be blue.
(2) [A ball is drawn from Jar B]. The ball is likely to be blue.
(3) [A ball is drawn from Jar A]. If the ball had been drawn from B, it would have been likely to be blue.

In 3 experiments varying red/blue proportions (total \(N = 1400\)), we found near-perfect correlations between responses to (1-3) with likely, might, have to, certain, and no modal (all \(r’s > .95\)). This implies a unified interpretation.

Core puzzle. Consider a probabilistic semantics based on \([Y07,Y10,Y12]\). \(\text{likely} \ P(\phi) > .5\), where \(i = (s,P)\) is a set of worlds and a probability measure. This accounts for (2): if \(P\) encodes that the ball is from B, \(P(\text{blue}) = .7\), which exceeds .5.

[Y12] extend the restrictor theory [L75,K86] to probabilistic indicatives: If \(\chi, \phi \text{ is likely}\) has \(L\ F(\phi) = 1\) iff \(P(\phi | O) > .5\). Semantically, the antecedent restricts \(i\) for the purpose of evaluating the consequent: \(P(\chi, \phi) = 1\) iff \(P(\phi | O) > .5\), where \(\chi\) restricts \(i\) to \(i' = (s \in \chi \ P(\chi) = 1\). This accounts for (1): if you know nothing about where the ball came from, \(P(\text{blue} | B) = .7\).

The challenge is (3), which seems perfectly parallel to (1-2). We would like to interpret (3) as having LF \(\chi\phi\) [\text{would likely}] [if \(\text{were B})\] blue, and associate it with an interpretation along the lines of “\(P_{\text{CF}}(\text{blue} | B) > .5\)”. But what is \(P_{\text{CF}}\)—what does “counterfactual probability” mean?

Causal relevance. Whatever CF probability is, it had better be sensitive to causal relations (cf. [B99,K01,E04]). Consider: Fran misses her flight; the pilot has a heart attack, and the plane crashes, killing 70% of passengers. Al says: “If F. had been aboard, it’s likely she would have died.” Case 1: F. is an ordinary civilian. Al’s claim seems true here. Case 2: F. has gone through pilot training, and probably could have landed the plane safely herself. Now Al’s claim seems false.

Causality interacts differently with indicative probabilities. Consider: we don’t know if F. made her flight. We do know that the plane crashed, killing 70%. Al says: “If F. was aboard, it’s likely she died.” Knowing that the plane crashed, this is true regardless of her piloting skills.

Conditioning and intervening. We can account for these puzzles by associating the indicative/CF divide with the distinction between conditioning and intervening [P00]. The measure \(P\) is derived from a structured causal model of the world: a Bayes net \(B\). We relativize interpretation to \(B\) and some observations \(O\): so, \(\text{likely} \ P(\phi | O) = 1\) iff \(P(\phi | O) > .5\), where \(P\) is the measure determined by \(B\). Bayes nets determine (conditional) probabilities by decomposing them into local probabilistic causal relationships among variables (“issues” [L88], “questions” [R96]). This Bayes net is a simple model of Case 1, where \(C\) (Crash?) is determined by \(H\) (Heart Attack?) but not by \(A\) (Aboard?). Letting \(D\) abbreviate Dead?, we find \(P(D)\) as \(\sum_X P(D | X) \times P(X)\), where \(X\) ranges over Boolean combinations of values for Parents(D) = \{A?, C\}. Since \(C\) has low prior probability, F. probably didn’t die: \(P(D) \approx .0017\).

Conditioning involves fixing variable(s) to values(s), allowing information to flow to causes and effects. If \(O = \{H,C\}\)—we know there was a crash, but we don’t know if F. was aboard—the relevant measure is \(P(D | O)\), which is \(\approx .35\) in this model. Our proposal
treats indicatives as adding an additional condition: against the informational background in \( O \), \( \text{If } F \text{ was aboard, it’s likely she died} \) holds iff \( P(D \mid C \land A) > 0.5 \). Given the independencies enforced by the graph, this equivalent to \( P(D \mid C \land A) \), which is given by the model as \( 0.7 \). So the probabilistic indicative is true in Case 1, as per intuition.

**Intervening** (Pearl’s “do” operator) sets a variable \( V \) to a value, severs connections to parents, and removes from \( O \) any observations involving \( V \) or its descendants. In the CF scenarios \( O = \{ H, C, A, D \} \)—heart attack, crash, F. not aboard, F. did not die. Al’s claim \( \text{If she had been aboard it’s likely she would have died} \) is true iff \( P(D \mid O, do(A)) > 0.5 \). We compute this by severing \( A \) from its parents (vacuously, in this model), removing \( A \) and its descendant \( D \) from \( O \), and conditioning on \( A \). This is equivalent to finding \( P(D \mid H \land C \land A) \) which—given independencies in the model—equals \( P(D \mid C \land A) = 0.7 \). This exceeds \( 0.5 \), so Al’s utterance is true in Case 1, as desired.

We see the importance of **causal relevance** by considering the contrast with **Case 2**. This model (see below) is identical other than the additional link from \( A \) to \( C \)—encoding the fact that F.’s presence might have prevented the crash—which necessitates a more complex conditional probability table for \( C \). Everything else remains the same, including \( O = \{ H, C, A, D \} \). To compute \( P(D \mid O, do(A)) \) we remove observation \( A \) and downstream \( C \) and \( D \), then condition on \( A \). The result is the equivalent to \( P(D \mid H \land A) \): the probability of death when heroic Fran was aboard and probably managed to prevent the crash. This corresponds closely to the intuitive evaluation, and the numerical computation yields \( \approx 0.176 \), well below \( 0.5 \). Because \( C \) is now a descendant of \( A \), Al’s utterance can be false in Case 2.

As noted above, moving to the **Case 2** model does not affect indicative judgments. The reason is that conditioning does not allow us to ignore the downstream observation that there was a crash. With \( O = \{ H, C \} \) as in the first indicative example, we compute \( P(D \mid O, A) \), holding fixed that \( C \in O \). As in the earlier indicative case, we end up with \( P(D \mid H \land A \land C) = P(D \mid A \land C) = 0.7 \). The indicative remains true in Case 2.

**Probabilistic triples.** This theory predicts the experimental result that (1-3) have identical acceptability varying probabilistic information. Consider a model where \( P(Jar = B) = 0.5 \). \( P(\text{blue} \mid B) \) is the proportion of blue balls in \( B \). (1) is true, against the specified background information, iff \( P(\text{blue} \mid O = \emptyset, B) > 0.5 \). This is equivalent to \( P(\text{blue} \mid B) > 0.5 \). (2) is true in its context iff \( P(\text{blue} \mid O = \{B\}) > 0.5 \), which reduces to the same truth-condition \( P(\text{blue} \mid B) > 0.5 \). (3) is true iff \( P(\text{blue} \mid O = \{A\}, do(B)) > 0.5 \). Removing \( A \) from \( O \) and conditioning on \( B \), we arrive at the same result. We thus have an explanatory account of why (1-3) pattern together empirically under so many conditions, which generalizes to other probabilistic language—probable/-ly, certain(ly) and arguably possible/-ly, perhaps, might, must, etc.

**Bare conditionals.** Causality is similarly relevant to bare CFs [B99,K01,E04,S11]. Consider the scenarios above, but with “70% died” adjusted to “100%”. Judgments about \( \text{If } F \text{ had been aboard, she would have died} \) are influenced by whether F. is a pilot, but the corresponding indicative judgments are not. The fact that causation trumps similarity (cf. [F75]’s Nixon example) is a well-known headache for theories following [L73] (e.g., the extensive hand-waving in [L79, 467-72]).

Can the present theory account for bare CFs? Perhaps. Bare restrictor conditionals require a silent operator, often thought to be epistemic **MUST**. If **must** has a probabilistic interpretation—e.g., \([\text{must } \phi]^B = 1 \) iff \( P(\lceil \phi \rceil)^B \) exceeds a threshold \( \theta \)—we derive the prediction that **If \( \chi \), \( \phi \) requires \( P(\phi \mid O, \chi) > \theta \). The corresponding CF requires \( P(\phi \mid O, do(\chi)) > \theta \). Correct judgments about causal (ir)relevance in the scenario just described follow immediately. This result does not depend on
whether $\theta = 1$ [Y05] or may be lower [S06,L16]. It does depend on the plausibility of probabilistic must, which we will not defend here; but it may yield an indirect argument for the latter theory, since it provides a simple, independently motivated explanation of causality effects in bare conditionals.

Related work. The basic idea explored here is drawn directly from [P00], with substantial further inspiration from [S11], but it does not seem to have been integrated into formal semantics in quite this way previously. Aspects of Pearl’s framework have been adapted by [S11,K13,S16], but these accounts have stripped away the probabilistic component which is crucial for us, and have focused on non-(explicitly-)modal conditionals. [S11] suggests an interpretation of conditionals in terms of interventions on Bayes nets for both indicatives and CFs; the account given here crucially relies on an asymmetry indexed by the choice of indicative vs. subjunctive morphology.

References


Y05: Yalcin, Seth. 2005. A puzzle about epistemic modals. MITWPL.

