Proceedings of SALT 27: 525–546, 2017

Probabilistic language in indicative and counterfactual conditionals

Daniel Lassiter
Stanford University

Abstract This paper analyzes indicative and counterfactual conditionals that have in their consequents probability operators: *probable*, *likely*, *more likely than not*, *50% chance* and so on. I provide and motivate a unified compositional semantics for both kinds of probabilistic conditionals using a Kratzerian syntax for conditionals and a representation of information based on Causal Bayes Nets. On this account, the only difference between probabilistic indicatives and counterfactuals lies in the distinction between conditioning and intervening. This proposal explains why causal (ir)relevance is crucial for probabilistic counterfactuals, and why it plays no direct role in probabilistic indicatives. I conclude with some complexities related to the treatment of backtracking counterfactuals and subtleties revealed by probabilistic language in the revision procedure used to create counterfactual scenarios. In particular, I suggest that certain facts about the interaction between probability operators and counterfactuals may motivate the use of Structural Equation Models (Pearl 2000) rather than the more general formalism of Causal Bayes Nets.

Keywords: Indicative conditionals, counterfactual conditionals, epistemic modals, probability operators, causal models, Bayesian networks

1 Conditionals, probability operators, and time

Here is an attractively simple, but flawed, attempt to account for the relationship between two kinds of conditionals in terms of the flow of time. (This is related to, but simplified relative to, ideas discussed by Adams (1975); Slote (1978); Edgington (1995) among others.) In many cases, the future indicative (2a) and the retrospective counterfactual (2b) seem to pattern together in terms of truth and/or assertibility.

(1) a. [Spoken on Monday:] If Mary plays in the game Tuesday, we will win.

b. [Mary doesn't play, and we lose. Spoken on Wednesday:] If Mary had played in the game Tuesday, we would have won.

The same holds for overtly probabilistic conditionals:

(2) a. [Spoken on Monday:] If Mary plays on Tuesday, we will probably win.

b. [Mary doesn't play, and we lose. Spoken on Wednesday:] If Mary had played on Tuesday, we probably would have won.

The idea is roughly: for $t_0 < t_1 < t_2$, an indicative uttered at t_0 about an event at t_1 is true just in case the matched counterfactual is, uttered at t_2. This hints at an attractive theoretical reduction, and a useful heuristic for constructing counterfactual scenarios—"rewind" to just before the antecedent time, making necessary changes to render the antecedent true, and consider what happens when the modified history unfolds again. Call this the RRR heuristic—"Rewind, Revise, Re-run".

However, Slote (1978) describes an example due to Sidney Morgenbesser that makes trouble for these claims about (1)-(2) (see especially Barker 1998). This case is known in the literature as "Morgenbesser's coin". Here is a variant:

(3) [A fair coin will be flipped on Tuesday. On Monday you are offered a bet: pay $50 and win $100 if the coin comes up heads, but nothing if it's tails.]

a. [Spoken on Monday:] If you bet, you will win.

b. [You don't bet on Monday, and the coin flip on Tuesday comes up heads. Spoken on Wednesday:] If you had bet, you would have won.

(3a) and (3b) seem to differ, but we should be careful not to misdiagnose the difference. The counterfactual (3b), spoken on Wednesday, is plainly true: the coin came up heads, and so if you had bet on heads you would have won. What is the status of (3a)? Well, it is clearly not *assertible* on Monday: there was no reason to favor heads over tails then. But it is plausible that it is *true* on Monday nonetheless, since heads did come up. The speaker made a lucky guess, but lucky guesses can be correct. This example might still be problematic depending on your theory of tense (cf. Belnap & Green 1994; MacFarlane 2003). But it does not yet refute the thesis floated above about the indicative/counterfactual connection.

A variant involving probability operators does refute this thesis, though.

(4) a. [Monday:] If you bet, there's an exactly 50% probability that you'll win.

b. [Wednesday:] If you had bet, there's an (exactly) 50% probability that you'd have won.

(4a), as spoken on Monday, is true. (4b) is false, though: since the coin came up heads, there is a 100% probability on Wednesday that you'd have won if you'd bet.

Morgenbesser's coin also shows that the RRR heuristic does not accurately describe our interpretative procedure for counterfactuals. The problem is that the coin flip is a *random* process, and it happened *after* you declined the bet. Rewind to Monday, Revise so that you take the bet, and Re-run history with this modification. With probability .5, the coin comes up tails on Tuesday and you lose. So the RRR heuristic predicts (4b) to be true, when in fact it is false.

Barker (1998, 1999) gives a diagnosis of the crucial feature of Morgenbesser examples: the temporally later fact that is held fixed—here, that the coin flip came up heads—is **causally independent** of the antecedent. The outcome of the coin flip does not depend in any way on your bet. The suggestion, as I will develop it, is that the "Rewind, Revise, Re-run" heuristic is *almost* right. When evaluating a counterfactual you should not throw out *all* information about events posterior to the antecedent time. Instead, you selectively throw out information about the outcomes of events that depend causally on the antecedent, and keep information about events that are causally independent of the antecedent. This corresponds to an elaborated heuristic: Rewind to the antecedent time, Revise to make the antecedent true, and selectively Regenerate following events that depend causally on the antecedent.

Kaufmann (2001a,b), Edgington (2004, 2008), and Hiddleston (2005) take up Barker's suggestion and incorporate it into a semantics for bare counterfactuals in various ways. Edgington also discusses, informally, how to accommodate the explicitly probabilistic examples like (4) that Barker focuses on. Focusing on conditionals with overt probability operators, I will argue that these approaches are on the right track for counterfactuals, and that the modified heuristic just described has a precise and illuminating formal implementation in terms of *interventions* on causal Bayes nets ("CBNs"; Meek & Glymour 1994; Pearl 2000). I will also show that probabilistic indicatives behave differently from matched counterfactuals: modifying causal information does not affect indicatives in the same way. To explain this difference, the compositional semantics will interpret probabilistic indicatives and counterfactuals identically modulo the choice between conditioning and intervening. In the last section I show that the revised heuristic is still empirically inadequate due to the special way that probability operators import subjective information into truth-conditional meaning, and suggest that this problem can be used to motivate a variant of the theory that is based on Structural Equation Models (Pearl 2000).

Because of length constraints I will not, however, deal with one crucial topic: the relationship between probabilistic conditionals and bare conditionals. Somewhat paradoxically, the key analytic tools used here—the restrictor theory of conditionals, and causal Bayes nets—both render bare conditionals more complicated than those with overt epistemic operators. In future work I hope to compare a number of strategies for extending this account to bare conditionals, building on Kratzer 1981, 1991a; Stalnaker & Jeffrey 1994; Kaufmann 2005, and Pearl 2000.[1]

[1] While similarity-based and premise-semantic accounts have long been dominant in philosophy and linguistics, many theorists have discussed the idea that counterfactuals depend on causation rather than vice versa—too many to undertake a detailed comparison with all relevant previous work in this space, unfortunately. Here is a partial list of work not cited in the main text, focusing just on philosophy and linguistics: Jackson 1977; Kvart 1986, 1992; Schaffer 2004; Schulz 2007, 2011; Briggs 2012; Kaufmann 2013; Stalnaker 2015; Zhao 2015; Santorio 2016. If we included psychology

2 Causal relations are crucial—but only for counterfactuals

Causal relations are directly relevant to the interpretation of probabilistic counterfactuals, but not to probabilistic indicatives. In this section I will give empirical evidence for this claim and begin to flesh out the "Rewind, Revise, selectively Regenerate" heuristic described in §1 more precisely. §3 implements the revised heuristic formally, shows how it interacts with causal and probabilistic information to generate counterfactual probabilities, and explains why indicatives differ.

Consider the following scenario, which verifies the counterfactuals in (5). (This morbid tale is modified from Edgington 2004.)

> **Version 1**: On the way to the airport to fly to Paris, Fran's car gets a flat tire. She misses her flight. During the flight, the pilot has a heart attack, the plane crashes, and 70% of the passengers are killed.

(5) a. If Fran had made her flight, it is likely/probable that she would have died.

b. If Fran had made her flight, there is a 70% chance she would have died.

These are Morgenbesser counterfactuals: their truth depends on holding fixed two contingent propositions, **heart-attack** and **crash**, that occur after the time of the antecedent **made-the-flight**. Since **heart-attack** and **crash** are causally independent of whether Fran made her flight, the revised, causally sensitive heuristic—"Rewind, Revise, selectively Regenerate"—holds them fixed in the counterfactual scenario.

Consider now a slightly enriched version of the story:

> **Version 2**: Same as Version 1, except that Fran is a highly skilled pilot who could easily land a passenger jet safely.

In this context, the probabilistic counterfactuals in (5) seem to be false. The revised heuristic gives a sense of how this is possible. Once we add the information that Fran is a pilot, the crash is no longer causally independent of whether she is aboard. Instead, there is a causal connection between her presence and the crash—since she could have landed the plane safely, whether the crash occurs depends on both **heart-attack** and **made-the-flight**. The revised heuristic therefore instructs us to ignore the contingent information that the crash happened.

(both cognitive and social), computer science, and statistics the list would be much longer. I have been inspired by many of these sources, and have focused on the ones discussed in the main text primarily because they bring out some of the key issues that I am concerned with here—causal relevance and the interaction with the language of probability. In particular, the large majority of applications of causal models in formal semantics have focused on deterministic models, which are logically simpler but insufficient to account for the phenomena considered here.

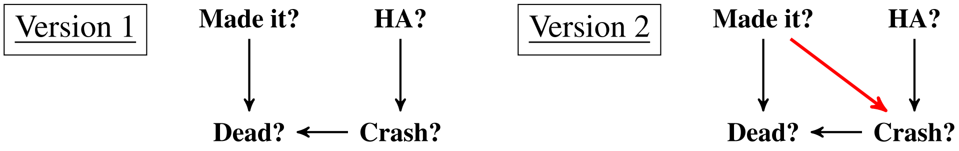


Figure 1 Causal structure of the two stories, with changes in version 2 in red.

As a step toward formalizing this reasoning, consider the simplified causal model of these scenarios in Fig. 1. Nodes pick out questions of interest (partitions on W, or "variables" in statistical jargon). Arrows indicate direct causal connections: $Q_1 \to Q_2$ indicates that the answer to Q_1 is an input to the function that determines the probability distribution on answers to Q_2. If there is a path $Q_1 \to \ldots \to Q_n$, we say that Q_1 is *upstream*/an *ancestor* of Q_n, and Q_n is *downstream* of Q_1.

In the causal model of version 1 on the left side of Figure 1, there is no path from **Made-it?** to either **HA?** ("Did the heart attack occur?") or **Crash?**. Instead, **Made-it?** is relevant only to the question **Dead?**: Fran presence and whether the crash occurs jointly determine whether she dies. This models the intuition that the heart attack and the plane crash are both causally independent of Fran's presence in the first version. The only variable that depends on whether she is aboard is whether she dies. In the second version the question of whether Fran makes the flight is causally relevant to the question of whether the plane crash occurs. Modeling this dependence requires us to add an arrow from **Made it?** to **Crash?**.

We can now state the revised heuristic a bit more precisely:

> When evaluating a counterfactual, (i) identify the question Q that the antecedent answers, (ii) discard the current answer to Q, (iii) add the information that the antecedent is true, and (iv) regenerate portions of the model that are downstream of Q, ignoring the actual answers to questions that are causally downstream of the antecedent. Do not change anything that is not causally downstream of the antecedent.

Note that the heuristic no longer refers to time. While the causal modeling approach does not require us to take a strong stance on the issue, it suggests the tantalizing possibility that the apparent relationship between counterfactuals and time may be a mere side effect of the fact that causal relations unfold over time—i.e., causes invariably precede their effects. The phenomena discussed in this paper seem to be consistent with this strong stance.[2]

[2] Step (i) hides a complexity: how do we proceed with complex antecedents, when there may not be a question that corresponds to the antecedent? I will not deal with this issue here, but see Briggs 2012; Ciardelli, Zhang & Champollion to appear; Lassiter 2017a for some starting points.

This heuristic handles the probabilistic Morgenbesser example (4) easily. The causal structure is just **bet?** → **win?** ← **heads?**, with the question of whether you won a joint effect of two causally unrelated events—whether you bet, and how the coin comes up. When revising to make **bet?** true, we throw out the downstream fact that **win?** is false, but keep the causally independent fact that **Heads?** is true. In the revised model, you bet and the coin comes up heads. *There is an exactly 50% chance that you win* is false in this counterfactual scenario: the true probability is 1.

We now have an explanation of why, in the flight scenario, intuitions about the sentences in (5) can be reversed by the addition of a causal connection between Fran's presence and the crash. In Version 1 (Figure 1, left), **Crash?** is causally independent of **Made it?**. When we revise to make **Made it?** true, we throw out the downstream observation that Fran is not dead (**Dead?** = F), but we hold fixed that the crash happened, since this question is causally independent of the antecedent. In a revised model with these features—there is a crash, and Fran is on board—we can make a prediction about the value of **Dead?** based on background knowledge about the causal structure of the scenario. The results are pretty grim: since 70% of the passengers died in the crash, presumably there is a 70% chance that Fran dies. This part of the reasoning is still informal, but we will make it precise in the next section.

The revised heuristic tells us to proceed differently in version 2, where Fran's presence on the flight is causally relevant to the crash (Figure 1, right). In this case, we throw out not only the fact that Fran is not dead, but also the fact that the crash happened, since both are downstream of the antecedent question **Made-it?**. The only question in the model whose answer we keep fixed is **HA?**—we hold fixed the causally independent fact that the pilot had a heart attack. In the revised scenario, Fran is aboard, the pilot is incapacitated, and we must use background causal knowledge to determine whether the crash happens and whether Fran dies. Based on the informal description given above, we may expect that Fran will save the day, the crash will not happen, and she will not die. (Again, we will see a formal model that derives this prediction in full in the next section.)

Manipulations of causal relevance lead to an interesting reversal of intuitions about the probabilistic counterfactuals in (5)—and we have the beginnings of a theory of counterfactual interpretation that makes sense of this reversal. What is the status of indicatives? Consider a variant of the original story, revised slightly to satisfy the felicity requirements of the indicative (i.e., that its antecedent's truth-value should not be known).

> **Version 1—unknown**: Fran was supposed to fly to Paris, but her car got a flat tire on the way to the airport. We don't know if she made the flight or not. We do know the pilot had a heart attack, the plane crashed, and 70% of the passengers were killed.

Against this background, consider the probabilistic indicatives corresponding to (5).

(6) a. If Fran made her flight, it is likely/probable that she died.

b. If Fran made her flight, there is a 70% chance that she died.

The sentences in (6) are, regrettably, true; we can only hope that Fran is not skilled at changing tires. Do intuitions reverse when we learn that Fran is a pilot?

> **Version 2—unknown**: Same as Version 1—unknown, except that Fran is a highly skilled pilot who could easily land a passenger jet.

No: once we know that the plane crashed, changing the story so that Fran is a potential hero does not make a difference to the probabilistic indicative. The fact of the crash is held fixed, even though it is causally downstream of the antecedent.

We will now formalize the interaction between conditional type and causal structure, with special attention to deriving the needed indicative and counterfactual probabilities while keeping the interpretation procedure maximally uniform.

3 Indicative vs. counterfactual as conditioning vs. intervening

3.1 Causal Bayes nets

Causal Bayes nets (CBNs) can be defined using familiar concepts from intensional and degree semantics. A causal Bayes net B is given by $\langle W, \mathcal{Q}, \mathcal{A}, \mathcal{C} \rangle$, where

- W is a set of possible worlds (the "sample space").
- $\mathcal{Q}$ is a set of questions ("variables"), where each $Q \in \mathcal{Q}$ is a partition on W.
- $\mathcal{A}$ is an acyclic binary relation on $\mathcal{Q}$ (the "arrows"). $\langle Q_i, Q_j \rangle \in \mathcal{A}$ means that Q_i is one of Q_j's parents, and has an immediate causal influence on Q_j.[3]
- A set of conditional probability tables $\mathcal{C}$, which determine a unique conditional probability distribution $P(Q_i \mid Parents(Q_i))$ for each question $Q_i \in \mathcal{Q}$.

I assume the standard definitions of (finitely additive) probability: P's domain is an algebra of subsets of W, $P(W) = 1$, and $P(X \cup Y) = P(X) + P(Y)$ whenever $X \cap Y = \varnothing$. The conditional probability $P(A \mid B)$ is $P(A \cap B)/P(B)$ whenever $P(B) \neq 0$, undefined otherwise. In addition, the probability measure P associated with Bayes net B is

[3] Note that this relation is automatically acyclic if we assume that causes precede effects and that temporal precedence is acyclic. Santorio (2016) considers a case which might motivate considering causal models with cyclic causal paths. If so, we can no longer think of the model as describing actual causation, as I have implicitly been assuming.

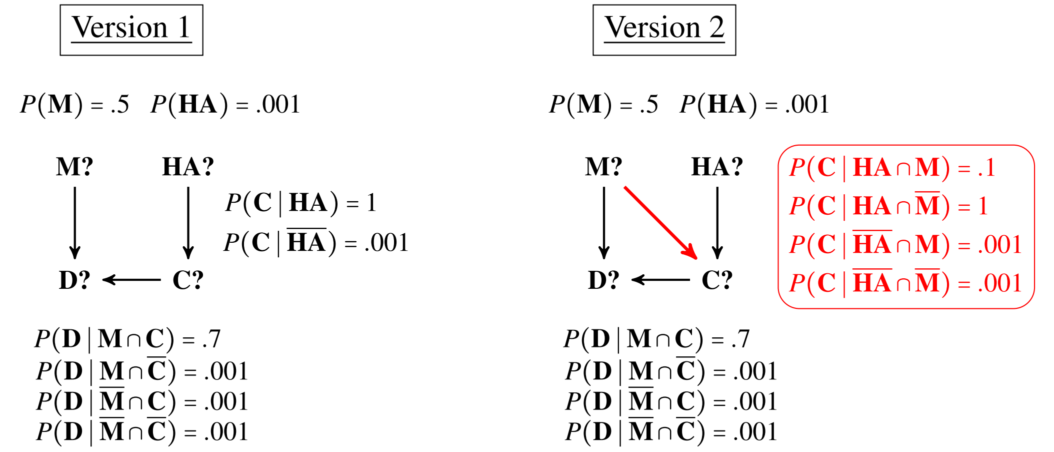

Version 1
$P(\mathbf{M}) = .5$ $P(\mathbf{HA}) = .001$
M?
HA?
D?
C?
$P(\mathbf{C} \mid \mathbf{HA}) = 1$
$P(\mathbf{C} \mid \overline{\mathbf{HA}}) = .001$
$P(\mathbf{D} \mid \mathbf{M} \cap \mathbf{C}) = .7$
$P(\mathbf{D} \mid \mathbf{M} \cap \overline{\mathbf{C}}) = .001$
$P(\mathbf{D} \mid \overline{\mathbf{M}} \cap \mathbf{C}) = .001$
$P(\mathbf{D} \mid \overline{\mathbf{M}} \cap \overline{\mathbf{C}}) = .001$
Version 2
$P(\mathbf{M}) = .5$ $P(\mathbf{HA}) = .001$
M?
HA?
D?
C?
$P(\mathbf{C} \mid \mathbf{HA} \cap \mathbf{M}) = .1$
$P(\mathbf{C} \mid \mathbf{HA} \cap \overline{\mathbf{M}}) = 1$
$P(\mathbf{C} \mid \overline{\mathbf{HA}} \cap \mathbf{M}) = .001$
$P(\mathbf{C} \mid \overline{\mathbf{HA}} \cap \overline{\mathbf{M}}) = .001$
$P(\mathbf{D} \mid \mathbf{M} \cap \mathbf{C}) = .7$
$P(\mathbf{D} \mid \mathbf{M} \cap \overline{\mathbf{C}}) = .001$
$P(\mathbf{D} \mid \overline{\mathbf{M}} \cap \mathbf{C}) = .001$
$P(\mathbf{D} \mid \overline{\mathbf{M}} \cap \overline{\mathbf{C}}) = .001$

Figure 2 Causal models including illustrative conditional probability tables specifying the distribution on each question conditional on its parents.

required to obey the *Markov condition*: each variable is probabilistically independent of its non-descendents, given the values of its parents.

A CBN, as defined here, does not generally determine a unique probability measure P. This is because the set of questions $\mathcal{Q}$ may provide only a very coarse grain on the set of possible worlds. However, we can simplify by assuming that there is one possible world for each cell in the maximally fine-grained question given by the intersection of the questions in $\mathcal{Q}$. For example, in the causal models in Figure 1 there were four questions, each with two answers (true/false). The simplification amounts to the assumption that there are at most 2^4 possible worlds—one for each combination of answers to the four questions. When this assumption is satisfied, each causal Bayes net B will determine a prior probability measure P_B.

For example, here are two CBNs that result from enriching the models in Figure 1 with conditional probability tables. (The conditional probability distributions used in Fig. 1 are meant to represent sensible assumptions, but are fairly arbitrary in the details.) Note that the distribution on **C?** in Version 1 depends directly only on **HA?**, and is (by the Markov condition) independent given a value of **C?** of any other variable except its daughter **D?**. In the model of Version 2, the addition of a causal relation between **M?** and **C?** requires a more complex model of the distribution on **C?**. In this case, the probability of **C?** depends on both **M?** and **HA?**.

These models allow us to answer many useful queries. For example, in Version 1 the probability of death, given that Fran made it and the pilot had a heart attack, is

$$\begin{aligned} P(\mathbf{D} \mid \mathbf{M} \cap \mathbf{HA}) &= P(\mathbf{D} \mid \mathbf{M} \cap \mathbf{C}) \times P(\mathbf{C} \mid \mathbf{HA}) + P(\mathbf{D} \mid \mathbf{M} \cap \overline{\mathbf{C}}) \times P(\overline{\mathbf{C}} \mid \mathbf{HA}) \\ &= .7 \times 1 + .001 \times 0 = .7 \end{aligned}$$

The first part of the equation—$P(D \mid M \cap HA) = P(D \mid M \cap C) \times P(C \mid HA)$—holds because the conditional independencies encoded in the graph allow you to decompose complex and long-distance queries into chains of simpler, more local queries. The rest is just look-up of values in the conditional probability tables. In the second model this same probability would decompose differently, due to the difference in causal structure: **C** depends on both **M** and **HA**.

$$\begin{aligned} P(\mathbf{D} \mid \mathbf{M} \cap \mathbf{HA}) &= P(\mathbf{D} \mid \mathbf{M} \cap \mathbf{C}) \times P(\mathbf{C} \mid \mathbf{M} \cap \mathbf{HA}) + P(\mathbf{D} \mid \mathbf{M} \cap \overline{\mathbf{C}}) \times P(\overline{\mathbf{C}} \mid \mathbf{M} \cap \mathbf{HA}) \\ &= .7 \times .1 + .001 \times .9 = .0709. \end{aligned}$$

The substantial difference between these conditional probabilities makes sense. In version 1, conditioning on Fran's presence or absence has no effect on the probability of a crash. In version 2, $P(\mathbf{C} \mid \mathbf{HA} \cap \mathbf{M}) \ll P(\mathbf{C} \mid \mathbf{HA} \cap \overline{\mathbf{M}})$: Fran's presence lowers the probability of crash in case of heart attack from 1 to .1. So, the probability of death **D**—which is an effect of the crash—is also much reduced.

3.2 Overtly probabilistic language

I will use a variant (and in some ways simplified, in others more complex) of a probabilistic domain semantics for *likely*, *probable/probably*, and *chance* (Yalcin 2007, 2010).[4] We relativize interpretation to a world w and a Bayes net B. Simplifying as discussed above, I assume that causal Bayes net B determines a unique probability measure P_B. In addition, we relativize to a set of observations $\mathcal{O}$. $\mathcal{O}$ may, but need not, be interpreted as the total information of some agent. (For instance, the background information in the first version of the story would be represented by $\mathcal{O} = \{\textbf{Made-it?} = F, \textbf{Crash?} = T, \textbf{Dead} = F\}$.) The inclusion of $\mathcal{O}$ is crucial because judgments about whether ϕ is likely should not depend only on the prior information about ϕ's likelihood that a CBN encodes—nor should it depend on the full totality of facts about the world, which would be appropriate only when we are modeling objective probabilities or the information of an omniscient agent.

(7) $[\![\text{likely}]\!]^{w,B,\mathcal{O}} = \lambda d_d \lambda q_{\langle s,t \rangle} \,.\, P_B(q \mid \mathcal{O}) > d$

Let's assume that the meaning of *likely* in the positive form is fixed to a threshold of .5 in every context. (This is not quite right, but close enough for present purposes.)

(8) $[\![(pos) \text{ likely}]\!]^{w,B,\mathcal{O}} = \lambda q_{\langle s,t \rangle} \,.\, P_B(q \mid \mathcal{O}) > .5$

The effect is that, for any ϕ, $[\![\phi \text{ is likely}]\!]^{w,P} = 1$ iff $P(\phi) > .5$. I will assume that the *It is probable that* ϕ and *Probably* ϕ are equivalent to *It is likely that* ϕ.

When a percentage expression binds *likely*'s degree argument, the result denotes a probability condition that depends on the percentage expression. For instance:

[4] See also Yalcin 2012; Swanson 2006, 2011, 2015; Lassiter 2010, 2015, 2017b; Moss 2015.

(9) $[\![\text{exactly 70\% likely}]\!]^{w,B,\mathcal{O}} = \lambda q_{\langle s,t\rangle} \,.\, P_B(q \mid \mathcal{O}) = 0.7$

As for *chance*, I will sidestep the interesting compositional puzzles presented by *There is a (n%) chance/probability that* ..., assuming simply that both are equivalent to *It is n% likely that* ϕ. I will also ignore important questions about whether there are different kinds of probability involved in the interpretation of these expressions.[5] Note, however, that it is possible to vary the interpretation between objective and subjective probability by varying the content of $\mathcal{O}$. To model "the subjective probability of ϕ is 0.7", we fill this parameter in with the observations that (say) some relevant individual or group has made. To model "the objective probability of ϕ at time t is 0.7", we use all of the actual values of variables whose answer has been determined at or before t. (With the semantics proposed below, the same trick can be used to model subjective vs. objective readings of indicative and counterfactual conditionals, as well as assessor-sensitivity if desired.)

For the moment, I will not commit to any claims about whether the meanings of other epistemic expressions also invoke probability (see Lassiter 2017b).

3.3 Probabilistic indicatives: Restriction as conditioning

I assume a restrictor syntax (Kratzer 1991a) for both indicative and counterfactual conditionals. In this approach the antecedent is interpreted as modifying the value of a contextual parameter temporarily, for the purpose of evaluating the consequent. If the parameter modified affects the interpretation of an operator *Op* in the consequent, the interpretation of the consequent will be affected as a result. Since the key examples treated here have overt epistemic operators in the consequent, I will not take sides in the debate about how to extend this analysis to bare conditionals, where there is no overt operator for the antecedent to influence.

For probabilistic indicatives, the target interpretation is one in which the probability measure referred to by the operator in the consequent is conditioned on the information in the antecedent (Yalcin 2007, 2012). This means, roughly, treating the antecedent as a "virtual observation" for the purpose of evaluating the consequent. We can implement this analysis very simply by adding the antecedent to the set of observations that the probability measure in the consequent is then conditioned on.

(10) $[\![\text{If } \phi,\ Op\ \psi]\!]^{w,B,\mathcal{O}} = [\![Op\ \psi]\!]^{w,B,\mathcal{O}^+}$, where $\mathcal{O}^+ = \mathcal{O} \cup \{[\![\phi]\!]^{w,B,\mathcal{O}}\}$.

Since the probability operators defined above are always conditioned on $\mathcal{O}$, the effect of (10) is to add the interpretation of the antecedent as an additional condition.

For example, in the indicative-friendly **unknown** variants of the flight story,

[5] For example, subjective vs. objective: see Ülkümen, Fox & Malle 2015; Lassiter 2017c for evidence that this distinction is relevant to the semantics of English probability expressions.

we do not know whether Fran made her flight, or whether she has died. So the observations are $\mathcal{O} = \{\textbf{heart-attack}, \textbf{crash}\}$. (10) tells us to analyze (6a) as follows:

(11) $[\![\text{If Fran made her flight, it is likely that she died}]\!]^{w,B,\mathcal{O}=\{\mathbf{HA},\mathbf{C}\}}$

$= [\![\text{it is likely that she died}]\!]^{w,B,\mathcal{O}=\{\mathbf{HA},\mathbf{C},\mathbf{M}\}}$

$= P_B(\mathbf{D} \mid \{\mathbf{HA},\mathbf{C},\mathbf{M}\}) > 0.5$ (by (7))

$= P_B(\mathbf{D} \mid \mathbf{M} \cap \mathbf{C}) > 0.5$ (independencies in *B*, Fig. 2)

$= 0.7 > 0.5$ (look-up in *B*, Fig. 2)

Note that it does not matter to this reasoning whether there is a causal link between **made-it?** and **crash?**. This is because the definition in (10) simply adds a proposition to $\mathcal{O}$, without regard for how this proposition is connected causally to others. This feature explains why manipulations of causal structure, holding other relevant facts about probabilities fixed, do not matter for indicatives.[6]

3.4 Probabilistic counterfactuals: Restriction as intervention

The proposed interpretation of counterfactuals is minimally different from that of the matched indicatives. The gross syntax of conditional sentences is the same, except that the morphological differences signal different ways to use the antecedent to modify the parameters of evaluation for the purpose of evaluating the consequent.

In the last section I proposed that indicative morphology in the antecedent is associated with conditioning—addition of the antecedent to the set of observations $\mathcal{O}$, with no change to the structure of the background causal model *B*. Subjunctive morphology, in contrast, is associated with the more complex operation of **intervention** (Meek & Glymour 1994; Pearl 2000), which makes separate modifications to the *B* and $\mathcal{O}$ parameters. My treatment is in the spirit of these previous accounts, but differs in that the implementation is tailored to make the compositional semantics simple, and is designed to emphasize the independence of the two kinds of modifications—revisions to $\mathcal{O}$ and to *B*. Here is a first pass, assuming that ϕ answers a particular question Q_ϕ in the definition of *B*. Note the close similarity to the rule for interpreting indicatives given in (10).

(12) $[\![\text{If were } \phi\text{, would } Op\ \psi]\!]^{w,B,\mathcal{O}} = [\![Op\ \psi]\!]^{w,B*,\mathcal{O}*}$, where

a. $B*$ is *B* possibly with the removal of some causal links (more below).

[6] It is possible that $P(\mathbf{D} \mid \mathbf{M} \cap \mathbf{C})$ could be different in the models of the two scenarios, with the result that (11) could vary in truth-value. This would imply that there is some probabilistically relevant difference between the two scenarios that has not been included in the story. (For example, Fran the pilot might know which parts of the plane are less dangerous in a crash.) This does not undermine the thesis pursued here. Estimating parameters and choosing which variables to model might well be affected by the addition of such information, but the fact of the crash is held fixed in this reasoning.

b. $\mathcal{O}*$ is $\mathcal{O}$ minus any answer to Q_ϕ or any of its causal descendants, plus ϕ.

The first condition (12a) is used to regulate the availability of backtracking interpretations of counterfactuals. The second condition (12b) implements the "Rewind, Revise, selectively Regenerate" heuristic motivated above. A key advantage of the semantics presented here, to my mind, is to maintain a clean separation between issues around (a) which pieces of factual information are retained and which are ignored when we create counterfactual scenarios, and (b) how and to what extent we can backtrack, inferring from counterfactual antecedents to features of their causal ancestors. We will return to both conditions for further discussion below; but we can already see how this accounts for Edgington's flight examples.

The precise spell-out of (12a) will not matter because the consequent is causally downstream of the antecedent. So let's simplify by assuming that $B*$ is just B. The target sentence is *If Fran had made her flight, it is likely that she would have died*. For both versions of the story, we are evaluating this sentence against the observational background $\mathcal{O} = \{\overline{\mathbf{M}}, \mathbf{HA}, \mathbf{C}, \overline{\mathbf{D}}\}$. For version 1 (Figure 2, left), where **C?** is causally independent of **M?**, (10) tells us to compute the value of *she probably died* relative to $B* = B$ and a modified observation set $\mathcal{O}*$. Here, $\mathcal{O}*$ is $\mathcal{O}$ minus $\overline{\mathbf{M}}$ (which answers the question addressed by the antecedent) and **D** (which is downstream of **M?**), and with the addition of the antecedent **M**. So, $\mathcal{O}* = \{\mathbf{HA}, \mathbf{C}, \mathbf{M}\}$ for Version 1.

(13) $[\![\text{If Fran had made her flight, it's likely she would have died}]\!]^{w,B,\mathcal{O}=\{\overline{\mathbf{M}},\mathbf{HA},\mathbf{C},\overline{\mathbf{D}}\}}$
$= [\![\text{it's likely she died}]\!]^{w,B,\mathcal{O}=\{\mathbf{HA},\mathbf{C},\mathbf{M}\}}$,

We can stop here: the second line is identical to the second line of (11)! So, of course, the counterfactual is true: the counterfactual probability of death given that she made her flight is .7.

The close connection between (11) and (13) illustrates a strong prediction of the present theory about the relationship between probabilistic indicatives and counterfactuals, reminiscent of the flawed efforts discussed in §1 to fix an indicative/counterfactual connection on the basis of temporal information. On the present account, holding causal structure fixed, a probabilistic counterfactual whose antecedent is known to be false should always pattern with the matched indicative, evaluated in a context where the observations are the same except that the value of the antecedent is unknown. In experimental work not reported here due to space limitations, I have verified this prediction for the items *likely, probably, might, certain*, and *have to*, across a variety of probabilistic contexts.

There are two key differences in Version 2: the conditional probability tables are different (see Figure 2), and step (12b) operates differently. The truth-conditions for Version 2 depend on $\mathcal{O}* = \mathcal{O} - \{\overline{\mathbf{M}}, \overline{\mathbf{D}}, \mathbf{C}\} \cup \{\mathbf{M}\}$, where the observation that there was a crash (**C**) has also been removed as being causally downstream of **M?**. This is enough to explain why the addition of a causal link between **M?** and **C?** has the effect that it does:

(14) $[\![\text{If Fran had made her flight, it's likely she would have died}]\!]^{w,B,\mathcal{O}=\{\overline{\mathbf{M}},\mathbf{HA},\mathbf{C},\overline{\mathbf{D}}\}}$
$= [\![\text{it's likely she died}]\!]^{w,B,\mathcal{O}=\{\mathbf{HA},\mathbf{M}\}}$,

where the *B* in question is the CBN in Figure 2, right. Here, the counterfactual probability of death is just .0701, as we saw in the last section. This is well below .5, and so the sentence is false.

4 Some complexities

4.1 Backtracking

On (12a): It has often been thought desirable to rule out interpretations of counterfactuals like (15) which involve reasoning from effects to causes.

(15) [Jar A has 2 blue and 8 red balls. Jar B has 8 red and 2 blue. The ball drawn was blue, but we don't know which jar the ball came from.]
If the ball were red, it would probably have come from Jar B.

In this scenario, the choice of jar determines the probability of each color, via the number of red vs. blue balls. A minimal causal model is **Jar** → **Color**. So, a true reading of (15) involves reasoning from effect **Color** to cause **Jar**, i.e., backtracking.

Lewis (1979), for instance, argues that backtracking readings should be false as a default, and only true on a "special" reading. However, backtracking readings are fully compatible with Lewis' semantic theory: their absence (or "specialness") is enforced by informal conditions on the similarity ordering. Theories of counterfactuals built around causal models have generally taken a stronger stance, ruling out backtracking as part of the definition of intervention (notably Pearl 2000; though see Hiddleston 2005; Lucas & Kemp 2015 for important exceptions.) This idea has an especially simple implementation in Pearl 2000, who defines interventions on a model *B* relative to a "surgically modified" model $B\star$. $B\star$ is identical to *B* except that all incoming links are removed from the intervention site (the antecedent, for counterfactuals). In the modified model used to interpret *If the ball were red, ...*, the question "Which color is the ball?" has no parents, and so no causes. As a result, after graph surgery adding information about **Color** cannot influence the now-independent variable **Jar**. For Pearl, an intervention to make the ball red conveys nothing about which jar the ball came from.

However, there is extensive experimental evidence that English speakers do perceive backtracking interpretations of counterfactuals (Sloman & Lagnado 2005; Rips 2010; Dehghani, Iliev & Kaufmann 2012; Gerstenberg, Bechlivanidis & Lagnado 2013; Lucas & Kemp 2015). In the recent unpublished experiment mentioned briefly above, I also tested the acceptability of (15) and related probabilistic conditionals.

Participants did not hesitate to interpret this sentence as being highly acceptable as a description of a visual scene corresponding to the context described. Indeed, it was just as acceptable as the matched indicative *If the ball is red, it probably came from B* in a context where the color is unknown. Here, reasoning from effect to cause is uncontroversially available.[7]

These results show that backtracking interpretations of probabilistic counterfactuals are robustly available. So, our theory must make room for these interpretations. At the same time, the theory should not make backtracking obligatory, given the many intuitive (e.g. Lewis 1979; Khoo 2016) and experimental (e.g., Sloman & Lagnado 2005) demonstrations suggesting that non-backtracking interpretations are often more accessible. In the approach pursued here, we can make room for both, without affecting the account of the separate issue of causal (in)dependence and its effects on revision. In brief, the idea is that "graph surgery" is more flexible than Pearl allows. In the modified graph $B*$ in (12a), the causal links that are removed do not have to be between the intervention site and its immediate parents, but may be further upstream. At the extremes, the modified graph can be the original B (with no links removed) or have all incoming links to the antecedent removed (Pearl's approach). In complex models, there will be many intermediate choices.

The resolution of this ambiguity is somehow context-dependent (Lewis 1979; Kaufmann 2013; Khoo 2016), and it may also depend on the presence of epistemic language in the counterfactual consequent. The invocation of "context" does not, of course, constitute a predictive theory of when and why backtracking readings are available; it only leaves room for them as needed. Hopefully future work will make available a precise statement of when and why backtracking interpretations are available, which can then be integrated into the present theory. (See Arregui 2005; Schulz 2007; Dehghani et al. 2012; Khoo 2016 for some promising directions.)

4.2 Revision

Condition (12b) is meant to enforce the revised heuristic from sections 1-2: evaluate counterfactuals by holding fixed causally independent facts while revising facts that are causally downstream. By comparison to (15), this aspect of counterfactual interpretation has been much less emphasized in the literature on causal theories of counterfactuals. However, it is just as crucial: as Goodman (1947) already made clear, the key problematic in a theory of counterfactuals is to work out which aspects of the real world to hold fixed and which to allow to vary. This aspect of the

[7] The same held of matched indicative/counterfactual pairs with *likely*, *might*, *have to*, and *certain*, varying systematically information about the contents of the jars. For (15), the average slider rating was 84 out of 100, as compared to 81/100 for the matched indicative. The experiment is reported in a longer version of this paper, but is suppressed here due to the length limit.

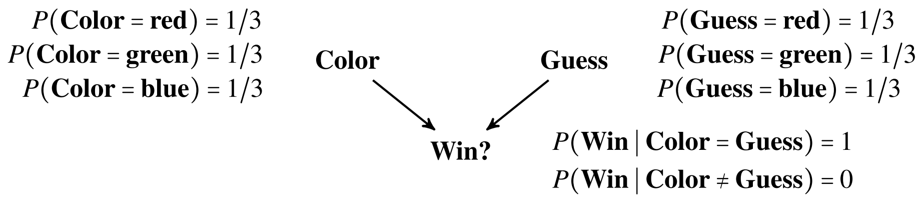


Figure 3 CBN for the guessing game.

interpretation of counterfactuals has been a primary focus of inquiry in work in both similarity-based theories (Lewis 1979) and premise semantics (Kratzer 1981, 1989; Veltman 1985, 2005, etc.). The revision condition (12b) is intended to play this role—and it seems possible to do without the additional formal machinery of similarity and premise sets, at least when treating probabilistic counterfactuals. This would be a useful reduction, since there is abundant independent evidence for the cognitive importance of probabilistic reasoning and causal models: see Glymour 2001; Sloman 2005; Gopnik & Schultz 2007; Griffiths, Kemp & Tenenbaum 2008; Tenenbaum, Kemp, Griffiths & Goodman 2011; Danks 2014 among many others.

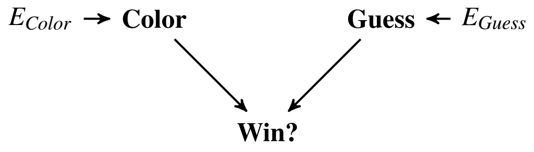
But the simplified statement in (12b) is, alas, too simple to capture the interplay between observational evidence, probabilities, and counterfactuals. Consider a guessing game: someone picks a ball at random from a jar containing equal numbers of red, blue, and green balls. You try to guess the color. If you get it right, you win. A CBN for this game is in Figure 4. Suppose that you guessed "blue". You are then told that you did not win, without learning which color did come up.

Against this background, the counterfactual in (16) seems to be true:

(16) If you'd guessed "red", there's an exactly 50% chance that you'd have won.

The trouble is that condition (12b) makes this counterfactual false. When evaluating the consequent of (16), (12b) tells us to ignore the information that I did not win, since it is causally downstream of the antecedent (which triggers an intervention on **guess**). However, the information that we did not actually win is needed here, because it is what allows us to infer that the ball was not blue. Conditioning on the fact that the ball is not blue, in turn, changes the probability of **Color = red** and **Color = green** from 1/3 to 1/2. An improved version of (12b) should allow us to incorporate into our model diagnostic information about the actual values of causally independent variables via conditioning, before we proceed to ignore their actual values when evaluating the effects of an intervention.

What we would like is for the probability distribution on variables causally independent of the antecedent to be informed by *all* relevant observations. In particular, the fact that you did not win should be considered, along with your actual



$$\forall C \in \{\textbf{red}, \textbf{green}, \textbf{blue}\} : P(E_{Color} = C) = 1/3$$
$$\forall C \in \{\textbf{red}, \textbf{green}, \textbf{blue}\} : P(E_{Guess} = C) = 1/3$$
$$\textbf{Guess} = E_{Guess}$$
$$\textbf{Color} = E_{Color}$$
$$\textbf{Win?} = (\textbf{Color} == \textbf{Guess})$$

Figure 4 Structural Equation Model for the guessing game.

guess, in deciding what the probability is that the color is red. So there should be some opportunity for information to flow from the real-world observation $\overline{\textbf{win}}$ to flow to **Color**. However, we obviously do not want the observation $\overline{\textbf{win}}$ to persist in the counterfactual scenario that we construct when evaluating (16). If it were, the probability of **win** would be 0, not .5, and the sentence would be trivially false.

While it may be possible to resolve this problem by complicating the definitions further, the simplest solution seems to be to switch to using the Structural Equation Models (SEMs) advocated by Pearl (2000). The key difference in these models is that, rather than including a conditional probability table for each variable V, we equip each V whose value is not a deterministic function of its parents with an exogenous source of randomness E_V. All causal relations in the model are deterministic, given values for these causeless variables.

If we formalize CBNs in this way, the definition in (12) can be modified to account for (16). The definition assumes that the consequent ψ does not make reference to any exogenous variables.

(17) $[\![\text{If were } \phi, \text{ would } Op\ \psi]\!]^{w,B,\mathcal{O}} = [\![Op\ \psi]\!]^{w,B*,\mathcal{O}*}$, where

a. $B*$ is a (possibly) surgically modified version of B, with all exogenous variables updated to reflect the information in the actual observation set $\mathcal{O}$.

b. $\mathcal{O}*$ is $\mathcal{O}$ minus any answer to Q_ϕ or any of its causal descendants, plus ϕ.

Again, since backtracking is not relevant here we can simply fix $B* = B$. Now $\mathcal{O} = \{\textbf{Guess} = \textbf{blue}, \textbf{Win?} = \textbf{False}\}$. Given these observations, the equations also allow us to infer (deterministically) that $\textbf{Color} \neq \textbf{blue}$, since otherwise **Win?** would be true. So, of course, $E_{Color} \neq \textbf{blue}$. Updating $P(E_{Color})$ with this information, we find that $P(E_{Color} = \textbf{red})$ and $P(E_{Color} = \textbf{green})$ are both .5. By (17a), this information about the exogenous variable E_{Color} is retained when we apply (17b), which tells us to ignore the observations **blue** and $\overline{\textbf{win}}$, leaving $\mathcal{O}* = \{\textbf{Guess} = \textbf{red}\}$. With probability .5, $E_{Color} = \textbf{red}$, and so $\textbf{Color} = \textbf{red}$; so, with probability .5 **Win?** is true. Equally, with probability .5, $E_{Color} = \textbf{green}$, and so $\textbf{Color} = \textbf{green}$ and **Win?** is false. As a result, (16) comes out true according to (17), as desired.

The key feature of (17) that makes this result possible is that exogenous random variables are treated specially: their probability distributions are updated on the basis of real-world observations $\mathcal{O}$, and not on the basis of the modified observations $\mathcal{O}*$ that are used elsewhere in the evaluation of the counterfactual consequent.

Finding a solution to the problem noted in this section has pushed us to shift from general CBNs to SEMs, and to use a semantics for counterfactuals very close to the "Twin Networks" approach of Pearl (2000: §7)—except, of course, that backtracking is not ruled out. Importantly, it seems to be the interaction between counterfactuals and probability operators that makes the problem visible. Probabilistic approaches to counterfactuals have generally assumed that the relevant values are derived from objective chances (see for example Edgington 1995, 2008; Bennett 2003; Kaufmann 2001a; Leitgeb 2012). In a semantics build around CBNs, these appear to correspond to the counterfactual probabilities that we would get from an accurate causal model of the world, if the true values of all variables were observed (in $\mathcal{O}$). If so, the problem discussed in this section would not arise, because it would not be necessary to use the observation $\overline{\textbf{win}}$ to infer the value of **Color?**: the true value of **Color?** is a fixed fact that is in $\mathcal{O}$. The problem may be apparent only when we consider the interaction with probability operators, for which the distinction between observed and unobserved variables is crucially relevant. The key motivation for shifting from general CBNs to SEMs can thus be seen to be due to special features of probability operators, and the way that they import subjective uncertainty into truth-conditional meaning.

This proposal also has important limitations. For example, it assumes that ϕ is not just a proposition, but also a complete answer to a unique question $\mathcal{Q}$ (value of a unique variable) in B. When these conditions are not satisfied—for example, when the antecedent is logically complex—the interpretation will have to be more complicated. This is a very general problem for theories of counterfactuals built around causal models, which are generally tailored for the case where each intervention is the assignment of a value to a variable, or a conjunction of such assignments. There is much work to be done here (see Briggs 2012; Ciardelli et al. to appear; Lassiter 2017a). I have also not specified what to do with conditionals with conditional (Kaufmann 2009) or epistemically modalized antecedents. If conditionals and epistemic sentences can be treated as denoting propositions, this problem may be reducible to the just-mentioned issue of identifying which interventions to perform in the case of antecedents that do not correspond to assignments of values to variables. But it remains to be seen whether this is feasible in general.

5 Conclusion

In this paper I have provided and motivated a unified compositional semantics for indicative and counterfactual conditionals that contain probabilistic language. I suggested that Barker's (1998) point that causal (in)dependence is fundamentally important for probabilistic counterfactuals—and the observation that the same does not hold for matched indicatives—can be accounted for by combining a restrictor approach to conditionals (Kratzer 1991a,b) with a representation of information using Causal Bayes Nets, which provide a simple format for the representation of causal and probabilistic dependencies. On this approach, probabilistic indicatives and counterfactuals are interpreted by using the information in the antecedent to modify the values of contextual parameters. The key difference between the two kinds of conditionals on this account is that indicatives monotonically add the antecedent as a virtual observation, while counterfactuals selectively remove observations, before adding the antecedent, in a way that is sensitive to the structure of the CBN. Optionally, counterfactuals may also break causal dependencies, limiting or prevent backtracking inferences by blocking the upward flow of information after conditioning on the antecedent.

In the final section, I pointed out a new theoretical problem for this account. When probability operators occur in counterfactuals, we need, paradoxically, both to ignore values of variables causally downstream of the antecedent, and to attend to them in order to infer the values of variables causally causally independent of the antecedent. I suggested that we can deal with this problem by shifting to the formalism of Structural Equation Models advocated Pearl (2000). The key property of these models is that uncertainty is represented using special exogenous variables, which can be treated differently from ordinary variables in the counterfactual interpretation procedure (as in (17)). While more work is clearly needed, this new puzzle may turn out to provide a reason to focus on SEMs in the analysis of counterfactual conditionals rather than general causal Bayes nets.

Daniel Lassiter
Stanford Linguistics
460 Margaret Jacks Hall
450 Serra Mall
Stanford, CA 94305
danlassiter@stanford.edu

References

Adams, Ernest W. 1975. *The Logic of Conditionals: An Application of Probability to Deductive Logic*. Springer.

Arregui, Ana Cristina. 2005. *On the accessibility of possible worlds: The role of tense and aspect*: University of Massachusetts PhD dissertation.

Barker, Stephen. 1998. Predetermination and tense probabilism. *Analysis* 58(4). 290–296.

Barker, Stephen. 1999. Counterfactuals, probabilistic counterfactuals and causation. *Mind* 108(431). 427–469.

Belnap, Nuel & Mitchell Green. 1994. Indeterminism and the thin red line. *Philosophical perspectives* 8. 365–388.

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

Briggs, Rachael. 2012. Interventionist counterfactuals. *Philosophical studies* 160(1). 139–166.

Ciardelli, Ivano, Linmin Zhang & Lucas Champollion. to appear. Two switches in the theory of counterfactuals: A study of truth conditionality and minimal change. *Linguistics and Philosophy* http://ling.auf.net/lingbuzz/003200.

Danks, David. 2014. *Unifying the Mind: Cognitive Representations as Graphical Models*. MIT Press.

Dehghani, Morteza, Rumen Iliev & Stefan Kaufmann. 2012. Causal explanation and fact mutability in counterfactual reasoning. *Mind & Language* 27(1). 55–85.

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

Edgington, Dorothy. 2004. Counterfactuals and the benefit of hindsight. In Phil Dowe & Paul Noordhof (eds.), *Cause and Chance: Causation in an Indeterministic World*, Routledge.

Edgington, Dorothy. 2008. Counterfactuals. In *Proceedings of the Aristotelian Society* 108 1, 1–21.

Gerstenberg, Tobias, Christos Bechlivanidis & David A Lagnado. 2013. Back on track: Backtracking in counterfactual reasoning, .

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

Goodman, Nelson. 1947. The problem of counterfactual conditionals. *The Journal of Philosophy* 44(5). 113–128.

Gopnik, Alison & Laura Schultz (eds.). 2007. *Causal Learning: Psychology, Philosophy, and Computation*. Oxford University Press.

Griffiths, Thomas L., Charles Kemp & Joshua B. Tenenbaum. 2008. Bayesian models of cognition. In Ron Sun (ed.), *Cambridge Handbook of Computational Psychology*, 59–100. Cambridge University Press.

Hiddleston, Eric. 2005. A causal theory of counterfactuals. *Noûs* 39(4). 632–657.

Jackson, Frank. 1977. A causal theory of counterfactuals. *Australasian Journal of Philosophy* 55(1). 3–21.

Kaufmann, Stefan. 2001a. *Aspects of the meaning and use of conditionals*: Stanford PhD dissertation.

Kaufmann, Stefan. 2001b. Tense probabilism properly conceived. In *Thirteenth*

Amsterdam Colloquium, 132–137.

Kaufmann, Stefan. 2005. Conditional predictions: A probabilistic account. *Linguistics and Philosophy* 28(2). 181–231. doi:10.1007/s10988-005-3731-9.

Kaufmann, Stefan. 2009. Conditionals right and left: Probabilities for the whole family. *Journal of Philosophical Logic* 38(1). 1–53. doi:10.1007/s10992-008-9088-0.

Kaufmann, Stefan. 2013. Causal premise semantics. *Cognitive science* 37(6). 1136–1170.

Khoo, Justin. 2016. Backtracking counterfactuals revisited. *Mind* .

Kratzer, Angelika. 1981. The notional category of modality. In Eikmeyer & Rieser (eds.), *Words, Worlds, and Contexts*, 38–74. de Gruyter.

Kratzer, Angelika. 1989. An investigation of the lumps of thought. *Linguistics and philosophy* 12(5). 607–653.

Kratzer, Angelika. 1991a. Conditionals. In A. von Stechow & D. Wunderlich (eds.), *Semantik: Ein internationales Handbuch der zeitgenössischen Forschung*, 651–656. Walter de Gruyter.

Kratzer, Angelika. 1991b. Modality. In Arnim von Stechow & Dieter Wunderlich (eds.), *Semantik: Ein internationales Handbuch der zeitgenössischen Forschung*, 639–650. Walter de Gruyter.

Kvart, Igal. 1986. *A theory of counterfactuals*. Hackett.

Kvart, Igal. 1992. Counterfactuals. *Erkenntnis* 36(2). 139–179.

Lassiter, Daniel. 2010. Gradable epistemic modals, probability, and scale structure. In Nan Li & David Lutz (eds.), *Semantics & Linguistic Theory (SALT) 20*, 197–215. CLC Publications.

Lassiter, Daniel. 2015. Epistemic comparison, models of uncertainty, and the disjunction puzzle. *Journal of Semantics* 32(4). 649–684.

Lassiter, Daniel. 2017a. Complex antecedents and probabilities in causal counterfactuals. In Alexandre Cremers, Thom van Gessel & Floris Roelofsen (eds.), *21st Amsterdam Colloquium*, 45–54.

Lassiter, Daniel. 2017b. *Graded Modality*. Oxford University Press.

Lassiter, Daniel. 2017c. Talking about higher-order uncertainty. Ms., Stanford University.

Leitgeb, Hannes. 2012. A probabilistic semantics for counterfactuals. Part A. *The Review of Symbolic Logic* 5(1). 26–84.

Lewis, David. 1979. Scorekeeping in a language game. *Journal of Philosophical Logic* 8(1). 339–359.

Lucas, Christopher G. & Charles Kemp. 2015. An improved probabilistic account of counterfactual reasoning. *Psychological Review* 122(4). 700–734.

MacFarlane, John. 2003. Future contingents and relative truth. *The philosophical quarterly* 53(212). 321–336.

Meek, Christopher & Clark Glymour. 1994. Conditioning and intervening. *The British journal for the philosophy of science* 45. 1001–1021.

Moss, Sarah. 2015. On the semantics and pragmatics of epistemic vocabulary. *Semantics and Pragmatics* 8. 1–81.

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

Rips, L. 2010. Two causal theories of counterfactual conditionals. *Cognitive science* 34(2). 175–221.

Santorio, Paolo. 2016. Interventions in premise semantics. *Philosophers' Imprint* .

Schaffer, Jonathan. 2004. Counterfactuals, causal independence and conceptual circularity. *Analysis* 64(284). 299–308.

Schulz, Katrin. 2007. *Minimal models in semantics and pragmatics: Free choice, exhaustivity, and conditionals*: ILLC, University of Amsterdam PhD dissertation.

Schulz, Katrin. 2011. "If you'd wiggled A, then B would've changed": Causality and counterfactual conditionals. *Synthese* 179(2). 239–251.

Sloman, Steven A. 2005. *Causal Models: How We Think About the World and its Alternatives*. OUP.

Sloman, Steven A & David A Lagnado. 2005. Do we "do"? *Cognitive Science* 29(1). 5–39.

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

Stalnaker, Robert. 2015. Counterfactuals and humean reduction. *A Companion to David Lewis* 57. 411.

Stalnaker, Robert & Richard Jeffrey. 1994. Conditionals as random variables. In *Probability and conditionals: Belief revision and rational decision*, 31–46. Cambridge University Press.

Swanson, Eric. 2006. Interactions With Context. Ph.D. thesis, MIT.

Swanson, Eric. 2011. How not to theorize about the language of subjective uncertainty.

Swanson, Eric. 2015. The application of constraint semantics to the language of subjective uncertainty. *Journal of Philosophical Logic* 1–26.

Tenenbaum, Joshua B., Charles Kemp, Tom L. Griffiths & Noah D. Goodman. 2011. How to grow a mind: Statistics, structure, and abstraction. *Science* 331(6022). 1279–1285.

Ülkümen, Gülden, Craig R Fox & Bertram F Malle. 2015. Two dimensions of subjective uncertainty: Clues from natural language. To appear in *Journal of Experimental Psychology: General*.

Veltman, Frank. 1985. *Logics for conditionals*: University of Amsterdam PhD dissertation.

Veltman, Frank. 2005. Making counterfactual assumptions. *Journal of Semantics*

22(2). 159–180.

Yalcin, Seth. 2007. Epistemic modals. *Mind* 116(464). 983–1026.

Yalcin, Seth. 2010. Probability operators. *Philosophy Compass* 5(11). 916–937.

Yalcin, Seth. 2012. Context probabilism. In M. Aloni, V. Kimmelman, F. Roelofsen, G. W. Sassoon, K. Schulz & M. Westera (eds.), *Logic, Language and Meaning* Lecture Notes in Computer Science 7218, 12–21. Springer.

Zhao, Michael. 2015. Intervention and the probabilities of indicative conditionals. *The Journal of Philosophy* 112(9). 477–503.