# Distinguishing semantics, pragmatics, and reasoning in the theory of conditionals

Daniel Lassiter
*University of Edinburgh*

## 1 Everything in its right place

In a crowded field of papers and books on conditionals, *The Meaning of If* (Khoo 2022) stands out as a major contribution. Philosophers and linguists will be studying and learning from it for many years to come. I find myself in agreement with much of the book's approach and tenor: philosophical concerns are balanced with careful linguistic analysis, against a set of background presuppositions derived from Bayesian epistemology/psychology, all in the service of deriving a comprehensive, unified picture of conditionals of all stripes. I also agree with many of Khoo's detailed claims about the semantics, pragmatics, and probabilities of conditionals, and the importance of causal models in the interpretation of subjunctive conditionals.

At the same time, reading the book was a bit like viewing my own understanding of conditionals in a funhouse mirror: the pieces are the same, but they are distributed in ways that I found jarring and difficult to rationalize. A comprehensive theory of conditionals has to account for how they are used in communication, reasoning, and deliberation, and it needs to respond to concerns from a wide variety of areas: a short list would include syntax, compositional semantics, Gricean pragmatics, speech act theory, the psychology of reasoning, epistemology, and decision theory. Khoo does an excellent job of balancing considerations from all of these fields, but the way that he incorporates them tends toward a sort of semantic imperialism. In his theory, the semantic component contains information and mechanisms that would, in my opinion, be better left to theories of how conditionals are used in communication and reasoning—pragmatics and psychology. In addition, various components of Khoo's theory are motivated exclusively by the need to account for facts about conditionals. One is naturally led to wonder whether there is a way to minimize ad hoc machinery, appealing as much as possible to independently motivated features of linguistic meaning and use, without compromising the theory's empirical coverage.

In what follows, I will compare Khoo's theory to a certain alternative picture that maintains a clean separation between the semantics of conditionals, the pragmatics of their use, and the psychological mechanisms involved in reasoning about uncertainty. By attending to interactions among these independently motivated components, we may be able to capture much of what is attractive about Khoo's picture of conditionals, while minimizing special-purpose semantic machinery and keeping everything in its right place.

## 2 Must we mix up facts and beliefs in the semantics of conditionals?

In chapters 3-4, Khoo motivates and formalizes his basic picture of the semantics and pragmatics of conditionals. Khoo is committed to Strong Centering: If $A$ is true at a world $w$, then "If $A$, $C$" has the truth-value of $C$ at that world, no matter what anyone believes. At the same time, Khoo's main purpose in ch.3 is to motivate the claim a conditional encodes a "constraint on inferential dispositions"—specifically, "the disposition to infer its consequent from its domain together with its antecedent". As a result, in Khoo's semantics "conditional contents are *mixed*—providing a constraint both on inferential dispositions and factual information" (p.51). True-antecedent conditionals are ordinary factual claims; false-antecedent conditionals are a fairly complicated sort of doxastic operator, and are associated with an expressivist pragmatics.

On a technical level, this works roughly as follows (with inspiration from van Fraassen 1976; Stalnaker & Jeffrey 1994; Kaufmann 2004). Rather than interpreting sentences relative to a possible world, we interpret each sentence relative to an infinite sequence of worlds, the first element of which is the world of evaluation. For non-conditional sentences, the infinite tail of this sequence is inert, and the result is identical to the usual single-world set-up. The same holds for a conditional "If $A$, $C$" evaluated in a world $w$ where $A$ is true: then we just check whether $C$ is true in $w$, without consulting the infinite tail. The rest of the sequence comes into play exclusively in cases where the sentence is a conditional, and its antecedent is false at the world of evaluation. In these worlds, the conditional's truth-value depends on the tail of the sequence, which encodes an agent's inferential dispositions or "learning plan". The latter is not determined by anything in the world of evaluation (or, perhaps, by anything at all, cf. Yalcin 2011). In ch.5, Khoo shows how to define probabilities of conditionals over this semantics in a way that approximates Stalnaker's Thesis [$P(\text{"If } A, C\text{"}) = P(C \mid A)$] but avoids Lewis' triviality proof (Lewis 1976).

Could we maintain the main attractions of this theory—factual truth-conditions for true-antecedent conditionals, a Bayesian picture of uncertain inference, and Stalnaker's Thesis—without pushing "inferential dispositions" into the semantics? A straightforward and promising approach was developed outside the mainstream conversation on conditionals by a number of philosophers, logicians, psychologists, and computer scientists. The core is the trivalent semantics, proposed originally by de Finetti (1936) and subsequently rediscovered by a number of authors (including Reichenbach, Jeffrey, Belnap, and Adams). On this theory, a conditional with a true antecedent is true or false, depending on the truth-value of its consequent; otherwise, it is undefined. Variants of the theory differ in terms of how they handle undefined values in Boolean combinations and in complex conditionals, and in the definition of logical consequence: see Égré, Rossi & Sprenger 2021 for a careful discussion of the options. To be clear, de Finetti's theory is at best a starting point for a comprehensive theory of conditionals. It obscures important syntactic facts about conditionals—notably, the fact that *if* is not a binary connective—and it makes odd predictions about certain complex conditionals (see, e.g., Appendix A of Khoo's ch.5). But these technical problems have fairly straightforward solutions (e.g., Égré et al. 2021).

Other than the obvious virtue of simplicity, the main attractions of de Finetti's theory for current purposes are two. First, it maintains a clear separation between what conditionals *mean*, and how people *reason with* them. Probabilistic information stays in the psychological/epistemological component of the theory. Second, it is supported by a wide variety of experimental evidence. One of

de Finetti's motivations for proposing this semantics was to show—several decades before Adams and Stalnaker made this a topic in philosophy of language—how the equation $P(\text{“If } A, C\text{”}) = P(C \mid A)$ can hold. The approach is straightforward: for any sentence $S$ , $P(S)$ is the probability that $S$ is true divided by the probability that $S$ is defined (e.g., Milne 1997; Cantwell 2006; Rothschild 2014). This proposal nimbly circumvents Lewis' proof and a variety of other triviality proofs for Stalnaker's thesis (Lassiter 2020). Its ability to support Stalnaker's Thesis without further ado is an obvious virtue, since the evidence for this equation is overwhelming (e.g., Hadjichristidis, Stevenson, Over, Sloman, Evans & Feeney 2001; Evans & Over 2004; Douven & Verbrugge 2010).[1] The empirical evidence around Stalnaker's Thesis does not, however, distinguish between de Finetti's theory from Khoo's, whose restricted support for Stalnaker's Thesis is also empirically adequate. (See below for discussion of cases involving partition sensitivity, where Khoo's predictions diverge sharply.)

More relevant to the present comparison, then, is the evidence around the status of conditionals with false antecedents. Since both theories endorse Strong Centering, Khoo's semantics and the trivalent theory differ in only one place: the value of a conditional when uttered at a world where its antecedent is false. In the trivalent approach, the conditional is simply undefined at this world. For Khoo, the infinite tail comes into play, and the interpretation changes from making a factual claim to expressing an inferential disposition (relativized to an agent's beliefs, not determined by facts of the world).

Peter Wason, the godfather of modern psychology of reasoning, is another theorist who re-discovered de Finetti's semantics, calling it the "defective truth-table" (Wason 1966). In his most famous experiment—now called the "Wason selection task", and the subject of a huge literature in psychology—he asked participants to collect evidence to judge whether a certain conditional rule held of a certain arrangement of cards, and observed which cards they turned over. Given "If $A$, $C$", they almost never turned over cards where $A$ was false, even though this was an optimal strategy for assessing the accuracy of the rule. Wason interpreted this and related results as showing that participants think of false-antecedent conditionals as being irrelevant, rather than potentially true or false. Further evidence for this interpretation is provided by experiments on conditional bets. In general, when I bet on $S$, I win if $S$ turns out to be true, and lose if $S$ turns out false. But what happens if I bet on "If $A$, $C$"? Participants's judgments are unequivocal: If $A$ is true, I win if $C$ is true and lose if it is false. But if $A$ is false, the bet is called off with all funds returned (Politzer, Over & Baratgin 2010).[2] To be clear, giving false-antecedent conditionals a different semantic status is not the only way to account for these results. For example, Khoo's theory could be extended to handle them by stipulating that (a) participants do not consult non-factual aspects of their interpretation of conditionals when judging the applicability of a conditional to concrete cases, and (b) non-factual

1 Recently Stalnaker's Thesis has been called into question by Douven, Elqayam & Mirabile (2022), who found a poor experimental fit in conditionals whose antecedents and consequents are irrelevant to each other. However, subsequent experiments indicate that Douven et al.'s experimental measures capture the pragmatic incoherence of the specific materials used, and their main result is not replicated with irrelevance conditionals that are designed to be coherent (Lassiter & Li 2024).

2 Over & Evans (2003, 2024)—who call it the "suppositional theory"—provide extensive discussion of relevant experimental evidence. In addition, Santorio & Wellwood (2023) have recently shown experimentally that the probabilities of conjunctions and disjunctions of conditionals do not obey the standard inequalities $P(A \wedge B) \leq P(A) \leq P(A \vee B)$. As they discuss, this is incompatible with a Boolean interpretation of conditionals, but consistent with trivalent semantics.

aspects of meaning are irrelevant when judging bets. But the trivalent semantics provides a much more straightforward explanation of why participants persistently treat false-antecedent cases as irrelevant.

In current research in psychology of reasoning, the trivalent theory of conditionals is one of the most prominent theoretical approaches. Why have philosophers and linguists been so slow to take it seriously? After all, among the large class of theories that endorse Strong Centering, all of the action in theories of conditionals revolves around the status of false-antecedent conditionals. The paradoxes of the material conditional are due to its choice to make them true; since making them false is even worse, truth-functional theories are therefore hopeless, assuming bivalence. The distinctive features of Stalnaker's selection semantics kick in only when we look at false-antecedent conditionals; and so on. A philosophical attachment to bivalence has surely played a role in this dialectic. David Lewis, for example, was aware that Belnap's trivalent theory circumvented his triviality proof, but dismissed this point in a footnote (in a different paper!), saying only that the addition of a truth-value was an "exorbitant" cost (Lewis 1975). One might have thought that triviality was a much greater cost.

To my knowledge, the only serious effort to argue on empirical grounds that false-antececent conditionals have non-trivial truth-values is by Khoo & Mandelkern 2019. They discuss the examples

(1) a. If Patch is a rabbit, then she is a rodent.
    b. It's not the case that, if Patch is a rabbit, then she is a rodent.

The claim that (1a) "is false, even if Patch isn't a rabbit", while (1b) is true regardless of whether Patch is a rabbit. These judgments align well with Khoo's (2022) claims that conditionals "encode inferential dispositions": no one with basic knowledge of zoology would be disposed to infer "Patch is a rodent" from "Patch is a rabbit". But the intuitive judgments are equally well explained by a simple pragmatic account. As usual, the sentences in (1) can only be appropriately asserted by someone who considers it possible, but not certain, that Patch is a rabbit. Appropriate assertion then requires (at a minimum) that the speaker believes these sentences. No one with basic knowledge of zoology could do so, and as a result (1a) is infelicitous in any normal context, and (1b) is potentially assertible in any normal context. We do not need to assume that these sentences have truth-values at worlds where Patch is not a rabbit. To account for the intuitions around these examples, we only need to think about the pragmatic demands on a speaker who lives at such a world, but does not know what species Patch is.

Khoo (2022), like many theorists before him, has the intuition that "If $A$, $C$" tells us something about what the relevant evidence would support, under the supposition that $A$ is true. This intuition is undoubtedly correct, but it is a rather large leap from here to the conclusion that conditionals are a sort of covert doxastic operator. The (correct) claim that conditionals "express inferential dispositions" is an observation about how conditionals are *used* in communication—a pragmatic issue, not a semantic one. One could certainly account for this by encoding inferential dispositions directly in the semantics of conditionals, supplemented by a story about when expressions of such dispositions are assertible. But a cleaner solution is to appeal to the pragmatics of assertion. If I tell you "It's sunny out", my evidence had better support my claim. This means (at a minimum, assuming I'm rational, etc.) that I should be in a state of information in which *It's sunny out* has

very high probability, conditional on my evidence. In a trivalent theory of conditionals, the same pragmatics tells us that a speaker who asserts "If $A$, $C$" had better be in a state of information in which $P(C \mid A)$ is very high, where $P$ is again implicitly conditioned on the speaker's evidence. But this is essentially the same as the speaker being disposed to infer $C$ from $A$—the non-factual part of the conditional's semantics, in Khoo's theory. The trivalent theory gets to the same basic destination, as far as the use of conditionals in communication is concerned; but it does so by decomposing it into a very simple factual semantics, and an independently motivated pragmatic theory. Probabilistic information remains in the psychological/epistemological component, rather than being duplicated in the semantics and mixed up with factual information.

## 3 Bounding puzzles

*The Meaning of If* starts with two "bounding puzzles". The first involves the observation that the update effect of a bare indicative conditional "If $A$, $C$" is the same as the update effect of the material conditional $A \supset C$. On most theories—including Khoo's—the indicative conditional asymmetrically entails the material conditional. On a standard interpretation of Stalnaker's (1978) pragmatics, this means that we should expect the update effect of "If $A$, $C$" to be stronger than that of $A \supset C$—it should eliminate more worlds. Khoo argues convincingly that it doesn't.[3] So, we have a paradoxical mismatch between the semantics of indicative conditionals and their pragmatic effects.

Khoo's solution to the the first bounding puzzle starts with a sophisticated empirical argument to the effect that that "factual beliefs fully determine inferential dispositions" (p.62ff). Chapter 4 (pp.98ff) uses this observation to motivate a correspondingly sophisticated pragmatics that involves a pair of operators, "Coarsening" ($\Downarrow$) and "Refinement" ($\Uparrow$). The interplay of these two operators manages mismatches between the "coarse" contents demanded by conversational update and the "refined" contents demanded by the semantics of conditionals. This footwork is needed in Khoo's theory because the infinite tail in the sequence semantics is able to make many distinctions that are required by need to "encode inferential dispositions", but are manifestly irrelevant to the conversational updates that these conditionals generate. The effect of the coarsening operator is to instruct the pragmatics to ignore these unwanted semantic distinctions. Khoo shows that the result is to render update with the indicative conditional equivalent to update with the material conditional, even though the indicative is sometimes false at sequences where the material conditional is true.

The solution works, but it involves a good deal of sophisticated pragmatic machinery that is tailored specifically to manage the update effect of conditionals. Can we achieve the same effect without any of this? In a trivalent semantics, the obvious update rule is the same as in Stalnaker's (1978) original proposal: eliminate worlds in which the sentence is false. For an indicative conditional "If $A$, $C$", this means eliminating only worlds where $A$ is true and $C$ is false. (Worlds where $A$ is false are let through not because they make the conditional true, but because they are treated as irrelevant.) But the $A \wedge \neg C$-worlds that are thereby eliminated are the same ones that are eliminated in update by $A \supset C$. This derives the first bounding puzzle without further ado. Khoo's first bounding puzzle is a real paradox, but only for bivalent theories: in trivalent theories, it

3 Khoo also argues later in the book that there are exceptions to this claim: see pp.169-70. Let's set this aside for a moment, as I'll return to it shortly.

is just what we would expect from de Finetti 1936 plus Stalnaker 1978. For the same reason, the trivalent semantics explains the *or*-to-*if* puzzle (Stalnaker 1975) in a maximally simple way.[4]

The second bounding puzzle involves subjunctive conditionals. Khoo argues convincingly that a subjunctive like (2a) is semantically weaker than the corresponding "strong subjunctive" (2b).

(2) a. If Alice had flipped the coin, it would have landed heads.

b. It couldn't have been the case that Alice flipped the coin, and it didn't land heads.

For instance, if we know the coin is fair and Alice didn't flip it, (2a) may or may not be true, but (2b) will be determinately false. Similarly, (2a) will have probability .5, but (2b) seems to have a probability of 0. So, (2a) does not entail (2b), and (2b) plausibly entails (2a). Nevertheless, "the information you learn from a subjunctive" like (2a) "seems to be that of a strong subjunctive" like (2b) (p.22). This puzzle thus pits the semantics of conditionals against their pragmatic effects again. If the update effect of an assertion of $S$ is to eliminate all worlds where $S$ is false, then (2a)—being semantically weaker—should admit some worlds that (2b) does not.

Khoo sets up intuitions around the subjunctive bounding puzzle with the following scenario (p.19). You know that the coin is either fair, double-headed, or double-tailed, and that Alice did not flip the coin. "[B]ut then you come to accept (perhaps via an assertion by a trusted informant)" that (2a) is true. You then conclude that the coin is double-headed. This is puzzling, since the truth of (2a) is compatible with the coin being fair. How did you rule out the worlds where the coin is fair, but (2a) is true?

The crucial part of Khoo's story is in the parentheses: "you come to accept *(perhaps via an assertion by a trusted informant)*". What do you learn when a person $x$ tells you $S$? As Egan & Sundell (2024) emphasize, what you learn in the first place is *not* that $S$ is true: it is that $x$ *asserted* $S$. A variety of further inferences may follow. Generally, you will also infer that $x$'s evidence supports $S$ (assuming $x$ is reliable, well-intentioned, etc.); that $x$ selected $S$ in an effort to balance considerations of informativity, relevance, politeness, and so on; and much more besides. Frequently, though by no means always, you will also be able to conclude that $S$ is true. (Compare also Grusdt, Lassiter & Franke 2022 on the complex interplay between pragmatic principles, background knowledge, and the update effects of indicative conditionals.)

So: how did your trusted informant $x$ come to have evidence supporting the truth of (2a)? If the coin is double-tailed, then $x$ obviously could not know (2a), so we can rule this out. If the coin is double-headed and $x$ knows it, then (2a) would seem to satisfy the major pragmatic desiderata, so this remains open. But if the coin is fair, and if the outcome of a fair coin flip is genuinely indeterministic, there is no way that $x$ could have come to know that (2a) is true *without seeing it flipped*. Since we have stipulated that the coin was not flipped, it follows that $x$ could not know that (2a) is true. You thus conclude that the coin is not fair—not by updating on the truth of (2a), but from the observation that $x$ asserted (2a) and the assumption that she was conforming to pragmatic norms.

This approach to the subjunctive bounding puzzle does not depend on any particular assumptions

4 Indeed, if we adopt the TT notion of consequence advocated by Égré et al. (2021)—which requires preservation of non-falsity—de Finetti's indicative conditional and the material conditional are mutually entailing, despite not being semantically equivalent (Égré et al. 2021: 197).

about the semantics of subjunctive conditionals. The same style of reasoning also problematizes Khoo's argument that the update effect of an indicative conditional "If $A$, $C$" is sometimes stronger than that of the material conditional $A \supset C$ (violating the "Weak Sufficiency*" condition: see p.169-70). In the "Edgington's Coin" scenario, Smith has a coin that is either double-headed or double-tailed, but we don't know which, or whether he flipped it. "Now suppose that Sue, *who has more information*" informs us of (3):

(3) If Smith flipped the coin, it landed heads.

I've added emphasis to the crucial bit here. To get intuitions going, we have to assume that Sue knows something we don't. If we do, we can get quickly to the conclusion that the coin is double-headed. This appears to violate our earlier conclusion that the update effect of "If $A$, $C$" is the same as that of $A \supset C$: after all, the material conditional is also compatible with the coin being double-tailed, as long as it's not flipped.

Here again, the puzzle is generated by the tacit assumption that the *only* thing we learn from an assertion of (3) is that (3) is true. Let's think further about what we can learn from the observation that Sue chose to assert (3), together with the assumption that Sue is better informed than us. What could Sue's mystery information be? She can't know whether or not the coin was flipped, or (3) would violate the open-antecedent condition of indicative conditionals. There are just two possible stories that remain. First, she might know that the coin is double-headed, without knowing whether it was flipped. Then, (3) would be true and supported by her information. Second, she might know the corresponding material conditional: that *either* the coin was not flipped, or that it was and that it landed heads. Then (3) would also be true and supported by Sue's information. However, it is difficult to reconstruct how Sue could have come by this complex disjunctive information, without knowing the truth-value of either disjunct. Without a plausible causal story about this process, a pragmatic interpretation of Sue's choice to utter (3) favors the simple explanation: Sue knows that the coin is double-headed.

This account predicts that apparent violations of Weak Sufficiency* should disappear if we modify the scenario so that there is a plausible causal story about how Sue could come to know the disjunction without any additional relevant evidence. This seems to be correct. For instance, suppose that Smith is playing the card game War with Jones, and Smith has just one card left—Jones has all the rest. Jones has turned over the card on top of his deck, which is an 8. Smith can either draw her top card, or give up and forfeit the game. (Her chances of winning are very low in any case.) Sue, who has been lurking outside the room, comes to you a moment later and says:

(4) If Smith turned over her last card, it was less than 8.

In this case, there is no temptation to take Sue's utterance to convey anything stronger than

(5) Either Smith didn't turn over her last card, or she did and it was less than 8.

Why don't we learn something stronger here—along the lines of "Smith's card was in fact less than 8", analogous to learning from (3) that the coin is double-headed? Because we can easily reconstruct how Sue could come to know this disjunction and nothing more, without inside knowledge: she observed an effect that counts equally as evidence for either disjunct. She might, for example, have seen the game conclude and the players leave, Jones smiling and Smith looking dejected. She

therefore surmises that Smith lost, and so (4) is true, without knowing what the card was or whether it was turned over.

The contrast between this scenario and Edgington's Coin is readily understood once we attend to the fact that the basic update effect of an assertion is not that the assertion is true: it is that the speaker has made the assertion (Egan & Sundell 2024). Updating with the information that the assertion is true will often—though not remotely always—be among the numerous effects of pragmatic reasoning about the basic fact of the assertion.

## 4 Partition sensitivity

In ch.6, Khoo argues that conditionals are semantically sensitive to partitions/Questions Under Discussion (QUDs), and that this fact is reflected in context-sensitivity in our intuitions about the probabilities of conditionals. The empirical arguments to this effect follow related observations of Kaufmann (2004) and Moss (2015). Partition-sensitivity is a major point of divergence between Khoo's theory and the trivalent semantics, which is firmly attached to Stalnaker's Thesis. For those who are attracted to the austere de Finettian theory, then, it is incumbent to develop an alternative explanation of why QUDs sometimes influence our intuitions about the probabilities of conditionals.

To illustrate, consider a morbid but useful example that Khoo adapts from Moss (2015).

> Jill was one of three people poised to jump from their respective buildings. You know that two of the three people had nets under them, and two of the three of them jumped. You know that one of the jumpers lived, that jumping without a net results in death, while jumping with a net results in survival. Thus, you know that one of the two who jumped did so with a net underneath them.

Assume we know nothing further about these people or the situation. Now consider the conditional

(6) If Jill jumped, she died.

The key observation is that our intuitions about the probability of (6)—at least sometimes, for some people—vary depending on what the Question Under Discussion is. If we are thinking of a question like "What happened if Jill jumped?", then a plausible response is that (6) has probability 1/2. After all, if she jumped, she was one of two who did; one of those people had a net and survived, and the other died. Since we have no reason to think that she was one or the other of these people, the chance that she died, given that she jumped, is 1/2.

However, we can also read (6) as a response to a less obvious question: "Did Jill have a net under her?" Then we can reason as follows (Khoo 2022: 192). Two of the three had a net under them. So, there's a 2/3 chance that Jill had a net under her. If she did and she jumped, she lived; otherwise she died. So, there's a 2/3 chance that (6) is false, and a 1/3 chance that it is true.

The former judgment (1/2) is what Stalnaker's Thesis predicts: the latter, responding to a less obvious but still plausible QUD, is inconsistent with Stalnaker's Thesis. I agree with Khoo that we should not treat the Thesis-violating judgment as a mere reasoning error along the lines of base-rate neglect, as Douven (2008) does. Khoo's response is to build QUD-sensitivity directly into the semantics of conditionals. While this does the job at hand, the addition of semantic partition-sensitivity to Khoo's theory adds further degrees of freedom to an already fairly complex theory.

Given that QUDs already play a large role in pragmatic theorizing in general—see, among many others, Beaver & Clark 2008; Ginzburg 2012; Roberts 2012; Kao, Wu, Bergen & Goodman 2014; Beaver, Roberts, Simons & Tonhauser 2017—one is led to wonder we could give a simpler and more explanatory account of partition-sensitivity in such pragmatic terms, rather than importing partitions into the semantics of conditionals.

Here is a sketch of how we might do so. Philosophers and psychologists alike have generally assumed that judgments of truth and probability are based on literal meaning, as theorized in compositional semantics. However, research in experimental pragmatics has show that this assumption is not entirely correct. When asked to judge the truth or probability of a sentence, many participants make a judgment about the message they think the sentence is intended to convey in context—complete with pragmatic enrichment. For instance, a classic experiment by Bott & Noveck (2004) found that around half of participants judged that the sentence *Some elephants are mammals* is false, apparently because it generates a false quantity implicature. When Cremers, Križ & Chemla (2017) designed a method to probe probability judgments for such "some" sentences, they found similarly that around half of participants reported the probability of the "some but not all" interpretation. That is, some participants reported the probability of the sentence's literal meaning; others reported the probability of the message that they thought the sentence was being used to convey, in context. Cremers et al. (2017) also found that a variety of other pragmatic factors influenced probability judgments, again with considerable variability in judgments.

Several other experiments corroborate the pragmatic sensitivity of probability judgment. For example, Fugard, Pfeifer, Mayerhofer & Kleiter (2010) examined probability judgments involving disjunction introduction in the consequent of a conditional. Imagining that a random number had been picked, they asked for probability judgments for the sentences in (7):

(7) a. If the number is 2, then it even.
    b. If the number is 2, then it is 2 or 4.

In terms of their literal meaning, both sentences should have probability 1. Most participants gave this response for (7a), but nearly all assigned probability 0 to (7b). Apparently, the pragmatic bizarreness of (7b) led participants to assign it a low probability, even though it is plainly true, because they could not recover a plausible message that it could be used to convey. In a different experiment, Lassiter & Li (2024) found evidence that failures of discourse coherence (Hobbs 1979; Kehler 2002) also influence probability judgments. They compared two types of conditionals whose antecedents and consequents were irrelevant to each other—ones where there was a recoverable coherence relation between clauses, and ones where there was not. Participants systematically judged the incoherent conditionals to have much lower probability than the corresponding conditional probability, but there was no similar effect for coherent conditionals. Here again, participants' probability judgments were strongly affected by what they took the conversational point of the sentence to be. If they could not recover one at all, they judged the sentence to have low probability.

My suggestion is that this is precisely what is going on in the cases that Kaufmann, Moss, and Khoo have used to motivate partition sensitivity. Admittedly, it would take much more ink and many more examples to spell out the case in convincing detail, but here is a quick sketch. When (6) is read as a response to the question "What happened if Jill jumped?", it is a direct answer to the question: "If she jumped, she died." No special pragmatics is needed in a typical conversation,

and we get a correspondingly ordinary probability judgment—the probability that (6) is true, in conformity with Stalnaker's Thesis.

In contrast, when (3) is read as an answer to the question "Was there a net?", it is not a direct answer—it is an *indirect speech act*, and the listener must do pragmatic work to figure out what the apparent non-answer is meant to convey about the QUD (Clark 1979). Consider the mini-conversation:

(8) a. **A**: Was there a net?
b. **B**: If Jill jumped, she died.

In the context we're considering, B's response clearly conveys a "No" answer to A's question: it is an indirect way to say that there was not a net. I suggest that we sometimes judge B's response to have probability 1/3 here because we are not evaluating the probability of its literal meaning. We are evaluating instead the probability of the message it conveys, in conversational context. If the relevant message is B's implied "No" answer to the question, the Thesis-violating response of 1/3 is what we should expect.

This new account of partition-sensitivity obviously needs to be spelled out in much more detail, and it must be confronted with the full range of examples that have been used to motivate this phenomenon. It does, for example, extend readily to other examples that Khoo discusses (e.g., the "McGee's Coins" example, pp.186-7). The indirect speech act theory of partition-sensitivity holds out the hope of the kind of theoretical simplification that Grice's (1975) "Modified Ockham's Razor" instructs us to search for. Rather than adding semantic complexity to the theory of conditionals, we should—wherever possible—try to account for the target phenomena in terms of independently motivated pragmatic and psycholinguistic phenomeona. The fully pragmatic approach also suggests an explanation of why our judgments about partition-sensitivity are fleeting and show a great deal of individual variability (Khoo 2022: 188). In this respect, they are just like the experimental results of Bott & Noveck (2004) and Cremers et al. (2017), where participants showed considerable variability in whether and how they incorporated pragmatic effects into their probability judgments.

## 5 Subjunctives, causal models, and tense

Chapters 7-10 of *The Meaning of If* turn to subjunctive conditionals. Khoo argues convincingly for the importance of causal models, incorporating a variant of the interventionist semantics built on causal models into his theory (e.g., Pearl 2000; Hiddleston 2005). Interventionist theories have long been popular in psychology, statistics, and computer science as well as philosophy of science; after a long period of neglect they have recently become prominent in linguistics and philosophy of language as well (Schulz 2011; Briggs 2012; Lassiter 2017; Ciardelli, Zhang & Champollion 2018; Santorio 2019). Khoo's choice here is, in my opinion, empirically and theoretically well-motivated. In ch.9, he also suggests a novel and promising way to modify interventionism to make room for backtracking readings of subjunctive conditionals.[5]

5 See, for example, Gerstenberg, Bechlivanidis & Lagnado 2013; Lucas & Kemp 2015 for some other ideas about how to extend the interventionist semantics to allow for limited backtracking. Khoo's approach, while presented as a challenge to the interventionist semantics, is compatible with other ways of implementing interventionism formally besides his preferred theory.

Khoo's path to these promising conclusions is fairly indirect. After an extended argument for the importance of tense in subjunctive reasoning, he notes that the straightforward implementation of the "past-as-past" approach that he advocates makes dramatically incorrect predictions about Morgenbesser examples (Slote 1978).[6] Many theorists have taken Morgenbesser cases to show that tense is not as important in a theory of subjunctive conditionals as it might have appeared. Interventionist theories, for example, are built directly around causal (in)dependence and make the correct predictions about Morgenbesser cases without further ado.

Khoo does not take this straightforward route. Instead, he maintains that the interpretation of subjunctive conditionals is built around *both* temporal and causal reasoning, and shows how to integrate the two by adding another powerful device to the semantics—ordering sources *à la* Kratzer 1981. He then stipulates that the ordering sources relevant to subjunctive conditionals include facts that are causally independent of the antecedent, while excluding those that are causally dependent (compare the very similar proposal of Kaufmann 2013). As a result, the refined domains of subjunctive conditionals privilege worlds where facts that are causally independent of the antecedent are unchanged.

The solution works, but it feels more like a technical trick than a genuine explanation. Ordering sources are a very powerful formal device, and there is no obvious reason why they should privilege causally independent facts over causally dependent facts. In other words, the addition of ordering sources to the theory allows Khoo to *simulate* the interventionist semantics, but it does not explain *why* the interpretation of subjunctive conditionals proceeds as it does. Without changing anything else, the theory could, for example, be modified to encode the opposite pattern or something completely orthogonal.

In contrast, a direct implementation of the interventionist semantics provides a constrained, explanatory theory of why causal dependence matters in subjunctive conditionals. On this approach, we evaluate a subjunctive conditional *If were A, would C* at a world $w$ by creating a modified world $w'$ in which $A$ is true, allowing this change to trickle down through the causal structure of the world without disturbing anything that is causally independent (Pearl 2000). (Think of it as a variant of the Stalnaker semantics, but one that applies only to subjunctives and provides a detailed algorithm for determining what the 'selected' world should look like.) Versions of the interventionist theory differ, for instance, in whether they consider a single way of intervening, or consider multiple interventions and quantify over them. But the key feature that unites these theories is that the role of causal dependence and independence in the semantics of subjunctive conditionals could not be otherwise:

6 These examples, first noted by Slote (1978), involve the observation that chancy facts that occur after the antecedent time, but are causally independent of the antecedent, are retained in subjunctive reasoning—in violation of the predictions of past-as-past. Assuming that a fair coin flip is genuinely indeterministic, the past-as-past theory makes the wrong prediction about (1):

(1) [Mary bet on tails, and then the coin was flipped. It came up heads.]
If Mary had bet on heads, she would have won.

In this context, (1) is clearly true, and has probability 1. But the past-as-past theory predicts that it should not be clearly true, and should have probability .5. If we return to the time when the antecedent was decided and change things so that Mary bets on heads, then the coin flip is in the future. When it happens again, there is a 50% chance that the coin will come up heads (so Mary wins) and a 50% chance that it will come up tails (so she loses).

it is baked deeply into the semantics of intervention. A direct implementation of interventionism thus explains *why* facts that are causally independent of the antecedent are unaffected, regardless of their temporal ordering. By contrast, the approach based on ordering sources makes it possible to encode this pattern, but it does not explain it.

There seem to be two major arguments in *The Meaning of If* for taking Khoo's circuitous route rather than the direct approach. The first involves the morphological connection between past tense and subjunctive marking in many (though not all) languages. As far as I can see, a proponent of the direct approach will have to bite the bullet here: there is no unified semantics for past tense and subjunctive morphology. I find this perspective less implausible than Khoo does (see his p.202f). In the interventionist theory, there is a close conceptual connection between subjunctive reasoning and time because causal relations unfold over time: if *A* causes *B*, then *A* temporally precedes *B*.

Across the languages of the world, there are many instances in which distinct morphosemantic functions are played by a single morpheme. Frequently, this seems to happen because there is either a conceptual connection between the two meanings or a substantial overlap in the contexts in which they are used. In either case, this situation favors linguistic changes in which a single morpheme can come to express two distinct functions which cannot be identified semantically. We might call this "non-accidental homophony" (though some instances may be better treated as cases of polysemy).[7] Even in English, the coincidence between tense and subjunctive morphology is not perfect: there are a number of distinct subjunctive forms, though the language is clearly in the final stages of a morphological reorganization involving the complete takeover of subjunctive marking by past tense.[8] This supports the suggestion that we are not dealing with two semantically identical categories, but two related categories which have a tendency to merge in their formal expression.

The second motivation for Khoo's roundabout way of encoding causal models is that it makes room for subjunctive conditionals to have an epistemic interpretation in some instances (pp.217ff). Khoo borrows an example from Edgington (1995): in a treasure hunt, the organizer says "It's either in the attic or the garden". I send my partner to search the attic, and go search the garden, finding

7 A few among many such examples are the tendency for languages to use the same morphology to encode imperfective and habitual meaning (Carlson 2012), and the coincidence of markers of topics and of conditional antecedents across many languages (Haiman 1978). While the latter has led some to argue for a theory of conditional semantics based entirely around topic-focus articulation, von Fintel (1994) and others have pointed out that this approach fails because conditional antecedents can be focused. However, if conditional antecedents are topical often enough, it is unsurprising from the perspective of language change that topic markers would frequently be re-analyzed in language change as morphological markers of the conditional antecedent. I suspect that the explanation of the connection between tense and subjunctive morphology is along the same lines: a strong but not universal pattern, motivated by robust tendencies of language change that are related to the conceptual organization of language and its use.

8 Older forms of English had a much more robust subjunctive, of which few remnants remain in modern Standard English. Examples are the first and third-person singular present-tense subjunctives of *to be*: *If I/he/she were there, ...*, and the use of *If I were to leave*, etc. Neither of these can be interpreted as past tense. The situation in modern English goes to show how messy and historically contingent generalizations about the morphological expression of such categories are. In informal spoken English and many non-standard dialects, the past tense forms are increasingly used instead of the historical subjunctives: *If I/he/she was*, corresponding to the imperfect *I/he/she was*. Likewise, the *were to* forms are moribund, and *If I were to leave* is increasingly perceived as an archaic variant of *If I left*. It appears that we are watching a change in progress involving the final demolition of the English subjunctive as a distinct category. But the fact that this demolition has happened gradually throughout the morphological system entails that there must have been a grammatical and conceptual distinction between tense and subjunctivity in the first place.

the treasure there. When my partner asks "Why did you tell me to search the attic?", I could readily reply:

(9) "Because if it hadn't been in the garden, it would have been in the attic: that's what I was told."

The problem for the interventionist theory is that there is no causal connection between the treasure's (not) being in the garden, and its being in the attic. As a result, if we intervene on the causal model implied by this scenario so that the treasure is not in the garden, there is no guarantee that it ends up in the attic. (Indeed, many interventionist theories do not even provide a clear story about what it would mean to intervene to make the treasure not-in-the-garden: see Lassiter 2017 for discussion and a proposal.)

I find this and Khoo's other examples of epistemic subjunctives convincing, and it is not yet clear whether a direct implementation of the interventionist semantics can account for them in a principled way. The best strategy for a proponent of the latter theory may be to treat the example in terms of backtracking. Specifically, the direct causal parent of the treasure's location is the organizer's choice about where to place it. If intervening to change the location of the treasure triggers backtracking to change the organizer's decision, and the only options available are the garden and the attic, then we may be able to account for why (9) seems true. I do not know whether this speculative suggestion will work out, but it may be the best hope for a slimmed-down interventionist theory to deal with such cases while relying only on independently motivated theoretical mechanisms.

Epistemic subjunctives, if they cannot be reduced to backtracking readings, seem to be a major point in favor of Khoo's way of encoding causal information in the semantics of subjunctive conditionals. Theorists interested in causation and counterfactuals would do well to attend carefully to Khoo's discussion of the issue in future work.

## 6 Conclusion

*The Meaning of If* is a carefully argued, comprehensive, and extremely stimulating contribution to the literature on conditionals. While I've focused here on points of disagreement, this does not detract from the overall importance of the book and the impressive range of ideas and concerns that it weaves together. My main plea is to make sure that everything stays in its right place: keep pragmatics and reasoning separate from the formal semantics. It's a fact of life for the working semanticist that we sometimes need complex semantic mechanisms and special-purpose pragmatic assumptions. But we should avoid these whenever possible, invoking them only when other avenues have been exhausted.

## References

Beaver, David & Brady Clark. 2008. *Sense and Sensitivity: How Focus Determines Meaning*. Wiley-Blackwell.

Beaver, David I, Craige Roberts, Mandy Simons & Judith Tonhauser. 2017. Questions under discussion: Where information structure meets projective content. *Annual Review of Linguistics* 3. 265–284.

Bott, Lewis & Ira A Noveck. 2004. Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of memory and language* 51(3). 437–457.
Briggs, Rachael. 2012. Interventionist counterfactuals. *Philosophical studies* 160(1). 139–166.
Cantwell, John. 2006. The laws of non-bivalent probability. *Logic and Logical Philosophy* 15(2). 163–171.
Carlson, Greg. 2012. Habitual and generic aspect. In Robert I. Binnick (ed.), *The oxford handbook of tense and aspect*, Oxford University Press.
Ciardelli, Ivano, Linmin Zhang & Lucas Champollion. 2018. Two switches in the theory of counterfactuals: A study of truth conditionality and minimal change. *Linguistics and Philosophy* 41(6). 577–621.
Clark, Herbert H. 1979. Responding to indirect speech acts. *Cognitive psychology* 11(4). 430–477.
Cremers, Alexandre, Manuel Križ & Emmanuel Chemla. 2017. Probability judgments of gappy sentences. In *Linguistic and psycholinguistic approaches on implicatures and presuppositions*, 111–150. Springer.
Douven, Igor. 2008. The evidential support theory of conditionals. *Synthese* 164(1). 19–44.
Douven, Igor, Shira Elqayam & Patricia Mirabile. 2022. Inference strength predicts the probability of conditionals better than conditional probability does. *Journal of Memory and Language* 123. 104302.
Douven, Igor & Sara Verbrugge. 2010. The Adams family. *Cognition* 117(3). 302–318.
Edgington, Dorothy. 1995. On conditionals. *Mind* 104(414). 235–329.
Egan, Andy & Timothy Sundell. 2024. All update is goat update. Ms., University of Kentucky and Rutgers University.
Égré, Paul, Lorenzo Rossi & Jan Sprenger. 2021. De Finettian logics of indicative conditionals part I: Trivalent semantics and validity. *Journal of Philosophical Logic* 50(2). 187–213.
Evans, Johathan St. B. T. & David E. Over. 2004. *If*. Oxford University Press.
de Finetti, Bruno. 1936. La logique de la probabilité. In *Actes du congrès international de philosophie scientifique*, vol. 4, 1–9. Hermann Editeurs Paris.
von Fintel, K. 1994. *Restrictions on quantifier domains*: University of Massachusetts at Amherst dissertation.
van Fraassen, Bas C. 1976. Probabilities of conditionals. In W.L. Harper & C.A. Hooker (eds.), *Foundations of probability theory, statistical inference, and statistical theories of science*, vol. 1, 261–308. Reidel.
Fugard, Andrew JB, Niki Pfeifer, Bastian Mayerhofer & Gernot D Kleiter. 2010. How people interpret conditionals: Shifts towards the conditional event. *Journal of Experimental Psychology: Learning, Memory, and Cognition* .
Gerstenberg, Tobias, Christos Bechlivanidis & David A Lagnado. 2013. Back on track: Backtracking in counterfactual reasoning, .
Ginzburg, Jonathan. 2012. *The interactive stance: Meaning for conversation*. Oxford University Press.
Grice, H. Paul. 1975. Logic and conversation. In P. Cole & J. Morgan (eds.), *Syntax and semantics 9: Pragmatics*, 41–58. Academic Press.
Grusdt, Britta, Daniel Lassiter & Michael Franke. 2022. Probabilistic modeling of rational commu-

nication with conditionals. *Semantics and Pragmatics* 1–59.
Hadjichristidis, Constantinos, Rosemary J Stevenson, David E Over, Steven A Sloman, JSBT Evans & Aidan Feeney. 2001. On the evaluation of 'If p then q' conditionals. In *Proceedings of the twenty-third annual conference of the Cognitive Science Society*, 409–414. Lawrence Erlbaum Associates.
Haiman, John. 1978. Conditionals are topics. *Language* 54(3). 564–589.
Hiddleston, Eric. 2005. A causal theory of counterfactuals. *Noûs* 39(4). 632–657.
Hobbs, Jerry R. 1979. Coherence and coreference. *Cognitive science* 3(1). 67–90.
Kao, Justine T, Jean Y Wu, Leon Bergen & Noah D Goodman. 2014. Nonliteral understanding of number words. *Proceedings of the National Academy of Sciences* 111(33). 12002–12007.
Kaufmann, Stefan. 2004. Conditioning against the grain. *Journal of Philosophical Logic* 33(6). 583–606.
Kaufmann, Stefan. 2013. Causal premise semantics. *Cognitive science* 37(6). 1136–1170.
Kehler, Andrew. 2002. *Coherence, reference, and the theory of grammar*. CSLI Publications.
Khoo, Justin. 2022. *The meaning of If*. Oxford University Press.
Khoo, Justin & Matthew Mandelkern. 2019. Triviality results and the relationship between logical and natural languages. *Mind* 128(510). 485–526.
Kratzer, Angelika. 1981. The notional category of modality. In Eikmeyer & Rieser (eds.), *Words, Worlds, and Contexts*, 38–74. de Gruyter.
Lassiter, Daniel. 2017. Complex antecedents and probabilities in causal counterfactuals. In Alexandre Cremers, Thom van Gessel & Floris Roelofsen (eds.), *Proceedings of the 21st amsterdam colloquium*, 45–54.
Lassiter, Daniel. 2020. What we can learn from how trivalent conditionals avoid triviality. *Inquiry* 63(9-10). 1087–1114.
Lassiter, Daniel & Chunan Li. 2024. Pragmatic intrusion in probability judgment: The case of conditionals. In *Proceedings of the 46th annual conference of the cognitive science society*, .
Lewis, David. 1975. Adverbs of quantification. In Edward L. Keenan (ed.), *Formal semantics of natural language*, 178–188. Cambridge University Press.
Lewis, David. 1976. Probabilities of conditionals and conditional probabilities. *Philosophical Review* 85(3). 297–315. doi:10.2307/2184045.
Lucas, Christopher G. & Charles Kemp. 2015. An improved probabilistic account of counterfactual reasoning. *Psychological Review* 122(4). 700–734.
Milne, Peter. 1997. Bruno de Finetti and the logic of conditional events. *The British Journal for the Philosophy of Science* 48(2). 195–232.
Moss, Sarah. 2015. On the semantics and pragmatics of epistemic vocabulary. *Semantics and Pragmatics* 8. 1–81.
Over, David E & Jonathan St BT Evans. 2003. The probability of conditionals: The psychological evidence. *Mind & Language* 18(4). 340–358.
Over, David E & Jonathan St BT Evans. 2024. Human reasoning. *Elements in Philosophy of Mind* .
Pearl, Judea. 2000. *Causality: Models, reasoning and inference*. Cambridge University Press.
Politzer, Guy, David E Over & Jean Baratgin. 2010. Betting on conditionals. *Thinking & Reasoning* 16(3). 172–197.

Roberts, Craige. 2012. Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics & Pragmatics* 5. 1–69.
Rothschild, Daniel. 2014. Capturing the relationship between conditionals and conditional probability with a trivalent semantics. *Journal of Applied Non-Classical Logics* 24(1-2). 144–152.
Santorio, Paolo. 2019. Interventions in premise semantics. *Philosophers' Imprint* .
Santorio, Paolo & Alexis Wellwood. 2023. Nonboolean conditionals. *Experiments in Linguistic Meaning* 2. 252–264.
Schulz, Katrin. 2011. "If you'd wiggled A, then B would've changed": Causality and counterfactual conditionals. *Synthese* 179(2). 239–251.
Slote, Michael A. 1978. Time in counterfactuals. *The Philosophical Review* 87(1). 3–27.
Stalnaker, Robert. 1975. Indicative conditionals. *Philosophia* 5(3). 269–286.
Stalnaker, Robert. 1978. Assertion. In Peter Cole (ed.), *Syntax and semantics 9: Pragmatics*, Academic Press.
Stalnaker, Robert & Richard Jeffrey. 1994. Conditionals as random variables. In *Probability and conditionals: Belief revision and rational decision*, 31–46. Cambridge University Press.
Wason, Peter C. 1966. Reasoning. In B. Foss (ed.), *New horizons in psychology*, 135–151. Penguin.
Yalcin, Seth. 2011. Nonfactualism about epistemic modality. In A. Egan & B. Weatherson (eds.), *Epistemic modality*, 295–332. Oxford University Press.