



Bernoulli Semantics and Ordinal Semantics for Conditionals

Stefan Kaufmann¹

Received: 28 June 2021 / Accepted: 16 May 2022

© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract

Conditionals with conditional constituents pose challenges for *the Thesis*, the idea that the probability of a conditional is the corresponding conditional probability. This note is concerned with two proposals for overcoming those challenges, both inspired by early work of van Fraassen: the *Bernoulli Semantics* associated with Stalnaker and Jeffrey, and augmented with a mechanism for obtaining “local probabilities” by Kaufmann; and a proposal by Bacon which I dub *Ordinal Semantics*. Despite differences in mathematical details and emphasis of presentation, both proposals lend themselves for use as a basis for a modal-theoretic interpretation of embedded conditionals.

The goal of this note is to compare the two frameworks by implementing a model for the interpretation of conditionals in each, based on the same underlying probability model for non-conditional sentences. I show that in the Ordinal model, certain sentences are assigned probabilities that do not accord with intuitions. This problem is familiar from the literature on Bernoulli models and can be addressed by introducing Kaufmann-style local probabilities into Ordinal models. I then show that Bernoulli Semantics has other limitations, in that it assigns probabilities in violation of the Thesis to certain very complex formulas. The upshot is that a fusion of the theories may be our best shot at getting the predictions right.

Keywords Conditionals · Probability · The Thesis · Bernoulli Model

1 Introduction

Compounded and embedded conditionals are an important test case for theories which assign probabilities to sentences in accordance with *the Thesis* – the idea that the probability of a conditional *if A, C* should equal the conditional probability of *C*

✉ Stefan Kaufmann
stefan.kaufmann@uconn.edu

¹ Department of Linguistics, University of Connecticut, 365 Fairfield Way, Unit 1145, Storrs, 06269, CT, USA

given A , whenever both are defined. The Thesis has been popular in the philosophical literature, and it enjoys strong and growing empirical support in psychology. Most of the discussions in these areas concern first-order conditionals (i.e., those whose constituents A and C do not contain conditionals). But what about sentences which do not fit this pattern? Such sentences are grammatically well-formed, easy to understand and widely attested, but the standard probabilistic calculus does not provide an obvious way to handle them. This is because the conditional probability of C given A cannot, on pain of triviality, be taken to be the probability of a proposition in the usual sense ([14, 25, 26], i.a.).

One way out of this predicament is to deny that truth conditions are the right vehicle for explicating the meaning and use of conditionals, either because they don't have truth conditions to begin with ([2, 5, 8], i.a.) or because their interpretation is subject to pragmatic constraints that end up doing all the work ([12, 15, 25], i.a.). An alternative is to maintain that truth conditions are crucially involved, but to rethink the relationship between truth conditions and possible worlds, at least as applied to conditionals.

The latter path has been explored in two directions, both inspired by van Fraassen's ([33]) seminal work. One direction, pursued by [17–21, 32] led to what I call the *Bernoulli* approach. In the other direction lay the *Ordinal* approach, proposed by [4]. The two differ in important respects, but share the same underlying intuition. It is not immediately clear whether either of them is more suitable than the other in building a semantic account of conditional sentences.

The goal of this paper is to examine this question. It is a straightforward exercise to define two models for conditionals – one Bernoulli and one Ordinal – based on the same underlying “simple” probability model for non-conditionals. When we do this, it turns out that the Ordinal model makes counterintuitive predictions for certain sentences; I will illustrate using an example Bacon mentioned (but did not discuss in detail). Kaufmann's version of Bernoulli semantics does not have that problem, and the same Kaufmann-style fix can be applied to the Ordinal model to bring its predictions in line with intuitions. On the other hand, there are cases (first discussed by [27]) where the Bernoulli model assigns probabilities that violate the Thesis. The Ordinal model is superior in these cases. The upshot is that a fusion of the two may ultimately be our best shot at getting the predictions right.

2 Background

I start with some key definitions.¹ To simplify matters, I will be using letters like ‘ A ’ interchangeably as variables over sentences and over propositions in the model. In all

¹For ease of comparison, I depart from the usual implementation of the Bernoulli account found in [32] and [18]. For one thing, in those works the denotation of a sentence is a random variable which may take values in the real interval $[0, 1]$. The Bernoulli model is merely an auxiliary tool for the calculation of those values. Here I omit this extra level and instead use the Bernoulli model directly for the interpretation. Secondly, in Kaufmann's version the theory was spelled out in terms of a sentential connective ‘ $\rightarrow$ ’ in the object language, representing the probability conditional. Bacon does not introduce an object-language sentential connective; instead, his conditional $\Rightarrow$ is a propositional operator defined on the measurable

such cases, the proposition A is intended to be the set of points (possible worlds or sequences thereof) at which the sentence A is true.

2.1 Simple probability models

Definition 1 (Simple probability model) A simple probability model is a structure $\langle W, \mathcal{F}, P \rangle$, where W is a non-empty set (of possible worlds); $\mathcal{F}$ is a σ -algebra over W ; and P is a probability measure on $\mathcal{F}$.

In this model, propositions are subsets of W ; proposition A is true at world w just in case $w \in A$. I assume throughout that atomic sentences are interpreted as denoting elements of $\mathcal{F}$, and that truth-functional connectives correspond to set-theoretic operations in the usual way. It has been known since [25] that in a simple probability model, except in certain degenerate cases, there is no binary propositional operator $\mathcal{O}$ such that the probability of $A\mathcal{O}B$ is guaranteed to be the conditional probability of B given A . Thus to satisfy the Thesis, we have to complicate the picture.

2.2 Bernoulli Semantics

In Bernoulli Semantics, the points in the probability space are not possible worlds but (countably) infinite sequences of possible worlds. I repurpose the term “proposition” to stand for sets of such sequences. Each sequence can be thought of as the outcome of a series of random selections from the set of possible worlds (with replacement), where the selections are mutually independent and uniformly governed by the probabilities in the simple model. I add a binary operator $\Rightarrow$ whose definition references the structure of the sequences: for first-order conditionals, $A \Rightarrow B$ is true at a sequence just in case B is true at the first A -world in that sequence (if any); the definition extends to more complex sentences, see below. There is a close conceptual relationship between this aspect of Bernoulli Semantics and Stalnaker’s ([29, 31]) Selection Semantics: the latter’s inspection of the *closest* A -world in case A is false at the world of evaluation, is replaced here with the idea of choosing an A -world *at random* (governed by the probability distribution).

Definition 2 (Bernoulli model) Let $\mathcal{M} = \langle W, \mathcal{F}, P \rangle$ be a simple probability model. The Bernoulli model derived from $\mathcal{M}$ is $\mathcal{M}^* = \langle W^*, \mathcal{F}^*, P^*, \Rightarrow \rangle$, where

- (i) W^* is the set of countably infinite sequences of worlds in W (allowing repetitions). Notation:
 - $w^*[n]$ is the n -th element of w^* ($n \geq 1$);
 - $w^*(n)$ is the tail of w^* that starts at $w^*[n]$.
- n.b. Note that for all w^* and n , $w^*(n) \in W^*$.

sets in his probability space. I follow this latter approach here, assuming that the English conditional is interpreted as the propositional operator $\Rightarrow$.

- (ii) $\mathcal{F}^*$ is the closure under complement and countable union of the *generating sets*, defined as follows:²

$$\{X_1 \times \dots \times X_n \times W^* \mid X_i \in \mathcal{F}, n \geq 1\}$$

- (iii) P^* is defined on the generating sets and extended to all of $\mathcal{F}^*$:

$$P^*(X_1 \times \dots \times X_n \times W^*) = P(X_1) \times \dots \times P(X_n)$$

- (iv) $\Rightarrow$ is a binary operator on $\mathcal{F}^*$, defined as follows:

$$X \Rightarrow Y = \{w^* \mid f^*(X, w^*) \in Y\}$$

Here f^* is a selection function defined as follows:

$$f^*(X, w^*) = \begin{cases} w^*(n) & \text{for the least } n \text{ such that } w^*(n) \in X \\ \lambda & \text{if there is no such } n \end{cases}$$

where λ is the “impossible” sequence at which all propositions are true.

There is some leeway in defining the conditional operator at sequences at which the antecedent X is not true anywhere. According to the above definition, the conditional is true at such sequences. This is the strategy of [32] and [33]. Kaufmann ([18, 19], and elsewhere) leaves the value of the conditional undefined in this case. The choice does not matter for present purposes: throughout this paper I assume that the antecedent has positive probability, thus the set of sequences at which it is not true anywhere has zero probability and can be ignored.

The following lemma, given here without proof, is helpful in calculating the probabilities of sentences containing conditionals.

Lemma 1 (Fraction Lemma – [33]).

If $P(A) > 0$, then $\sum_{n \in \mathbb{N}} P(\overline{A})^n = 1/P(A)$.

Let us say that a proposition has *rank 0* if it is of the form $A \times W^*$ for some $A \in \mathcal{F}$. In this case its probability is given by P in the simple model, since $P^*(A \times W^*) = P(A)$. Propositions not of rank 0 are of rank 1. The crucial difference between Bernoulli models and Ordinal models is that here there are no higher ranks.

The probability of a first-order conditional (both of whose constituents are of rank 0) can be calculated by assembling the proposition it denotes in $\mathcal{F}^*$ as a union of pairwise disjoint generating sets, each of whose probability under P^* is the product of probabilities under P . The next result establishes that the probabilities of first-order conditionals conform to the Thesis.

Proposition 2 (First-order conditionals).

If A, B are of rank 0 and $P(A) > 0$, then $P^*(A \Rightarrow B) = P(B|A)$.

²The closure conditions are mentioned in [32] but not in Kaufmann’s ([18–21]) presentations, although Kaufmann’s calculations obviously depended on them.

Proof

$$\begin{aligned}
 P^*(A \Rightarrow B) &= P^*\left(\bigcup_n (\bar{A}^n \times AB \times W^*) \cup \bar{A}^*\right) \\
 &= \sum_n P^*(\bar{A}^n \times AB \times W^*) + P^*(\bar{A}^*) \\
 &= \sum_n P(\bar{A})^n P(AB) + \lim_{n \rightarrow \infty} P(\bar{A})^n \\
 &= \frac{P(AB)}{P(A)} \quad \text{by the Fraction Lemma and since the second summand above is 0} \\
 &= P(B|A)
 \end{aligned}$$

□

As an aside, since $P^*(A \times W^*) = P(A)$ whenever A is in $\mathcal{F}$, it also follows that $P^*(A \Rightarrow B) = P^*(AB)/P^*(A)$. Thus we can say that in the Bernoulli model, $P^*(A \Rightarrow B)$ is both the probability of a proposition and the corresponding conditional probability.

The possibility of calculating the probability of a conditional in the Bernoulli model from probabilities in the underlying simple model, illustrated here for first-order conditionals, is available in general. For example, consider the case of a conditional $(A \Rightarrow B) \Rightarrow C$ where A, B, C are of rank 0, and assume that its antecedent $A \Rightarrow B$ has positive probability. The probability of this sentence is derived as follows (see also [18]).³

$$\begin{aligned}
 &P^*((A \Rightarrow B) \Rightarrow C) \tag{1} \\
 &= P^*\left(\bigcup_n (A \Rightarrow \bar{B})^n \times ((A \Rightarrow B) \cap C) \times W^*\right) \\
 &= P^*\left(\bigcup_n \left(\bigcup_m \bar{A}^m \times A \bar{B}\right)^n \times (ABC \cup \bigcup_m (\bar{A} C \times \bar{A}^m \times AB)) \times W^*\right) \\
 &= \sum_n \left(\sum_m P(\bar{A})^m P(A \bar{B})\right)^n \cdot \left(P(ABC) + P(\bar{A} C) \sum_m P(\bar{A})^m P(AB)\right) \\
 &= \sum_n P(\bar{B}|A)^n \cdot (P(ABC) + P(\bar{A} C) P(B|A)) \\
 &= \frac{P(ABC) + P(\bar{A} C) P(B|A)}{P(B|A)} \\
 &= P(C|AB)P(A) + P(\bar{A} C)
 \end{aligned}$$

The correspondence between P and P^* is straightforward: they coincide whenever both are defined, but the latter extends to propositions of higher rank. (Those higher-rank propositions “cut across” equivalence classes of sequences starting with the same world.) From now on, I will drop the typographical differentiation between the two whenever the distinction is not at issue.

³I omit the set of world sequences at which the antecedent is false throughout, i.e., the set $\overline{A \Rightarrow B}^* = \bar{A} \bar{B}^*$, which by assumption has zero probability.

2.3 Ordinal models

As already mentioned, Bacon's framework is also inspired by [33] and resembles the Bernoulli framework in key respects. In fact, every Bernoulli model is embedded in a corresponding Ordinal model, as will become clear momentarily. In this section I base the discussion on Bacon's "ur-conditional" operator $\Rightarrow$, for which the correspondance is the most immediate. Bacon also discusses a context-sensitive conditional operator that is indexed to an epistemic accessibility relation; I will say more on this below.

An ordinal model can be derived from a simple probability model $\langle W, \mathcal{F}, P \rangle$, in much the same way as the Bernoulli model above.⁴

Definition 3 (Ordinal model) Let a simple probability model $\mathcal{M} = \langle W, \mathcal{F}, P \rangle$ be given. The ordinal model derived from $\mathcal{M}$ is $\mathcal{M}_\infty = \langle W_\infty, \mathcal{F}_\infty, P_\infty, \Rightarrow \rangle$, where

(i) $W_\infty = W^{\omega_1} = \{\pi : \omega_1 \mapsto W\}$ is the set of ω_1 -sequences of worlds in W (ω_1 is the first uncountable ordinal)

- for $\alpha < \omega_1$, $W_\alpha = W^{\omega_\alpha}$ is the set of ω_α -sequences of worlds
- Since $W_\alpha \times W_\infty \cong W_\infty$ whenever $\alpha < \omega_1$, for each ω_α -sequence $\pi^\alpha \in W_\alpha$ there are sequences in W_∞ containing π^α as an initial sub-sequence followed by an element of W_∞ .

(ii) $\mathcal{F}_\infty$ is defined as follows (here ' $cl(X)$ ' is the closure of X under complement and countable union and intersection):

$$\mathcal{F}_0 = \{A \times W_\infty \mid A \in \mathcal{F}\}$$

$$\mathcal{F}_{\alpha+1} = cl \{A_0 \times \dots \times A_n \times W_\infty \mid A_i \times W_\infty \in \mathcal{F}_\alpha, 0 \leq i \leq n\}$$

$$\mathcal{F}_\gamma = cl \left(\bigcup_{\alpha < \gamma} \mathcal{F}_\alpha \right)$$

(iii) P_∞ is defined as follows:

$$P_0(A \times W_\infty) = P(A)$$

$$P_{\alpha+1}(A_0 \times \dots \times A_n \times W_\infty) = P_\alpha(A_0 \times W_\infty) \cdot \dots \cdot P_\alpha(A_n \times W_\infty);$$

extended to the rest of $\mathcal{F}_{\alpha+1}$ via
Carathéodory's theorem

$$P_\gamma(A) = P_\alpha(A) \text{ if } A \in \mathcal{F}_\alpha \text{ for } \alpha < \omega_1;$$

extended to the rest of $\mathcal{F}_\infty$ via
Carathéodory's theorem

(iv) $\Rightarrow$ is a binary operator such that for $A, B \in \mathcal{F}_\infty$:

$$A \Rightarrow B = \{\pi \mid f(A, \pi) \in B\}$$

⁴Bacon simplifies the exposition (without loss of generality) by assuming that W is countable and $\mathcal{F}$ is its powerset. I adopt these assumptions here.

Here f is a selection function defined as follows:

$$f(A, \pi) = \begin{cases} \pi[\omega^\alpha \cdot i] & \text{where } \alpha \text{ is the smallest number such that} \\ & A \in \mathcal{F}_\alpha \text{ and } i \text{ is the smallest number such that} \\ & \pi[\omega^\alpha \cdot i] \in A \\ \tau_A & \text{if there is no such number } i \text{ and } A \neq \emptyset \\ \# & \text{if } A = \emptyset \end{cases}$$

where τ_A is a designated (but arbitrarily chosen) A -sequence in W_∞ , and $\#$ is the “impossible” sequence at which all propositions are true.⁵

If we take a simple probability model $\langle W, \mathcal{F}, P \rangle$ as a common starting point, then the set of Bernoulli sequences W^* is W_1 in the ordinal model; similarly for $\mathcal{F}^*$ and $\mathcal{F}_1$. Bernoulli models have no more structure than that, whereas in Bacon’s models, W_1 is at the bottom of an infinite hierarchy of countable sequences of countable sequences of countable sequences of ... of worlds.

The definition of conditionals whose antecedent is of rank 0 is similar to that in the Bernoulli model. The intention behind the shift from π to $\pi[\omega^\alpha \cdot i]$ is again to skip over a finite (and possibly empty) initial sub-sequence of π until A is true for the first time. If A is a subset of W_0 , this means skipping worlds, as in the Bernoulli model. If π contains an A -world at some finite index i , the first line of the definition applies. The case where there is no such i is treated slightly differently here, differentiating between possible and impossible A ; but that difference is immaterial to the present discussion so long as we assume that the antecedent has positive probability.

The equality in (2) can be established by a parallel argument to that for Proposition 2 above.

$$P_\infty(A \Rightarrow B) = P(B|A) \quad (2)$$

With A of rank 0 the search for an A -world is restricted to a vanishingly small sliver of the sequence π , namely the first countable sub-sequence of worlds, corresponding to a complete sequence in the Bernoulli model. It is only with higher-rank antecedents that the rest of π becomes relevant.

Bacon’s main result for his ur-conditional, and the main selling point of his theory overall, is that Bayes’ Rule extends to conditionals of arbitrary (finite) ranks. Thus for instance, a derivation that is entirely parallel to (2) above, modulo the ranks of the constituents, applies to the following conditional with an antecedent of Rank 1:

$$Pr_\infty((A \Rightarrow B) \Rightarrow C) = Pr_\infty((A \Rightarrow B) \cap C) / Pr_\infty(A \Rightarrow B) \quad (3)$$

What I said above about the relationship between P^* and P also holds, mutatis mutandis, for P_∞ and P . I will omit the typographical distinction when there is no danger of confusion.

⁵Bacon calls α , the smallest number such that $A \in W_\alpha$, the “rank” of A . Since this is a semantic definition, there is no tight correspondence between the syntactic complexity of a sentence and the rank of its denotation. Therefore I will steer clear of using that notion in the discussion.

Sentences with the structure of (3) are particularly interesting here because the single one concrete example Bacon discusses at any length to motivate his use of Ordinal models is of this form. I will examine this case a bit more closely.

3 First-order conditional antecedents

While the result in (3) is reassuring, one may be forgiven for feeling somewhat unsatisfied. Whereas (2) above allowed us to ground the probability of $A \Rightarrow B$ in probabilities from the underlying simple probability model, no such link between P_∞ and P is established in the case of (3). In fairness, it was not Bacon's goal to provide such a link. But to deploy the framework in the interpretation of sentences relative to concrete probabilities, such a connection is needed. For both Bernoulli and Ordinal semantics, the underlying simple probability model determines the probabilities of (arbitrarily nested) conditionals, and one would want to use that fact to evaluate the resulting predictions. I will illustrate with an example.

Bacon discusses an English conditional of the form in (3) by way of arguing that such nested conditionals exist and ought to be in the scope of a semantic theory worth its salt. His discussion is informal, and he does not return to the example after the introduction of his formal apparatus. But we are given enough details (in the form of probabilities of atomic sentences and some conditional probabilities) to flesh out a model and see how it fares.

A derivation parallel to the one in (1) above shows that the probability of this sentence is the same as in the Bernoulli model:

$$P_\infty((A \Rightarrow B) \Rightarrow C) = P(C|AB)P(A) + P(\bar{A}C) \quad (4)$$

With that in mind, let us turn to Bacon's example sentence of this form, along with the intended background scenario (pp. 140-1) and his probability judgment:

Suppose you have ten numbered vases, three are shatter-proof and the remaining seven are fragile enough to break if dropped. You also know that two of the fragile vases are priceless, however, you don't know which of the vases are priceless or fragile. Suppose also that there has recently been an earthquake and there is a chance that some of the vases have fallen from their shelves onto the floor.

How confident should you be that vase number eight is priceless if it was one of the vases that broke if it was dropped? The intuitive answer is calculated as follows: there are seven vases that will break if dropped. Furthermore, we know that out of those only two are priceless, so the proportion of priceless vases out of those that broke if they were dropped is intuitively $2/7$.

The sentence in question is (5a). I give its denotation in (5b), where the variables stand for *Drop*, *Break*, and *Priceless*, respectively.

- a. If vase number eight broke if it was dropped, it is priceless.
 - b. $(D \Rightarrow B) \Rightarrow P$
- (5)

Bacon evidently assumes that any theory should predict a probability of $2/7$ for the sentence in the given scenario, but he does not show how to use an Ordinal model to derive predictions of this sort. However, the formula in (4) tells us what the predictions are. And in fact, it turns out that the probability we derive via this formula is not the desired $2/7$. This failure is instructive, for reasons already discussed in detail by [18]. Early versions of Bernoulli Semantics faced exactly the same challenge, and Kaufmann proposed an amendment which, modulo details of implementation, can also be used to adjust an Ordinal account.

Let us first see what the predicted probability is. In either framework, it is as given in (6).

$$P((D \Rightarrow B) \Rightarrow P) = P(P|DB)P(D) + P(\bar{D}P) \quad (6)$$

None of the three terms on the right-hand side of (6) is explicitly given in Bacon's description of the scenario, so we cannot immediately check whether it evaluates to Bacon's desired $2/7$. I will fill in the blanks as best I can by making what I take to be some very plausible independence assumptions; naturally I can only surmise that they are true to Bacon's intentions.

First, I assume that whether a vase is dropped is independent of whether it is priceless. Nothing in the story suggests otherwise. Thus the second summand can be factored as follows:

$$P(\bar{D}P) = P(\bar{D}) \times P(P) = P(\bar{D}) \times 2/10 \quad (7)$$

Furthermore, Bacon seems to assume that fragility is not a matter of degree: a vase that is dropped will break if and only if it is fragile. Thus in particular, $P(F|DB) = 1$ (where F stands for *Fragile*). Moreover, whether a fragile vase is priceless is conditionally independent of whether it is dropped, given its fragility. In other words, a fragile vase that is dropped (and breaks) is just as likely to be priceless as one that is not dropped; thus:⁶

$$P(P|DB) = P(P|DBF) = P(P|F) = 2/7 \quad (8)$$

Substituting (7) and (8) back into (6), we obtain (9). What this comes down to depends on $P(D)$, which is not given in the scenario; but since $P(D) + P(\bar{D}) = 1$, it is clear that the value must be strictly less than $2/7$.

$$P((D \Rightarrow B) \Rightarrow P) = 2/7 \times P(D) + 2/10 \times P(\bar{D}) < 2/7 \quad (9)$$

To be quite clear: the value derived here is the conditional probability of P given $D \Rightarrow B$, in accordance with the Thesis. Ordinal Semantics has done its advertised job. But the Thesis-obeying probability turns out to be a poor fit with the way people, including Bacon, interpret this conditional in the given scenario.

Consider again the way in which the antecedent $D \Rightarrow B$ is evaluated here. This is the (denotation of the) statement that if vase number eight was dropped, it broke. The model assigns this sentence the conditional probability $P(B|D)$, that an arbitrary vase broke given that it was dropped. This conditional probability is $7/10$. But in the

⁶Here are the intermediate steps: $P(P|DB) = P(PF|DB) + P(P\bar{F}|DB) = P(F|DB)P(P|DBF) + P(\bar{F}|DB)P(P|DB\bar{F}) = 1 \times P(P|DBF) + 0$.

scenario, none of the vases has a 7/10 conditional probability of breaking given that it is dropped.⁷ Instead, the vases fall into two classes: those that definitely break if dropped, and those that definitely don't. Vase eight is in one of these classes; we don't know which one. But we do know that which class a vase belongs to correlates perfectly with a property that is explicitly mentioned, and presumably causally independent of whether the vase is dropped: its fragility.

Kaufmann's ([18]) proposal, spelled out in the Bernoulli framework, is to bring such connections to bear on the evaluation of conditionals at sequences at which their antecedents are false. The original rule was to choose worlds at random repeatedly until one turns up at which the antecedent is true. The refined strategy is to continue choosing until a world turns up at which the antecedent is true *and relevant variables that are causally independent of the antecedent have the same value as they do at the base world* (i.e., the first world in the sequence). In terms of the resulting probabilities, this has the effect of restricting the domain of the random choices to those worlds which agree with the base world on the values of those causally independent variables. Kaufmann calls this a "local" interpretation of the conditional. The sentence's probability under a local interpretation obviously depends on the partition induced by the relevant variables; the ordinary, "global" conditional probability is a local interpretation relative to the trivial one-cell partition.

Consider a first-order conditional $D \Rightarrow B$ with a relevant variable F independent of D , whose possible values are F and $\bar{F}$. The local interpretation for this case is the following:

$$\begin{aligned} P(D \Rightarrow_F B) &= P \left(\frac{DB \times W^* \cup \bigcup_n (\bar{D}F \times \bar{D}\bar{F}^n \times DFB \times W^*) \cup \bigcup_n (\bar{D}\bar{F} \times \bar{D}\bar{F}^n \times D\bar{F}B \times W^*)}{\bigcup_n (\bar{D}F \times \bar{D}\bar{F}^n \times DFB \times W^*) \cup \bigcup_n (\bar{D}\bar{F} \times \bar{D}\bar{F}^n \times D\bar{F}B \times W^*)} \right) \\ &= P(DB) + P(B|DF)P(\bar{D}F) + P(B|D\bar{F})P(\bar{D}\bar{F}) \end{aligned} \quad (11)$$

which can be rewritten thus:⁸

$$= P(B|DF)P(F) + P(B|D\bar{F})P(\bar{F}) \quad (12)$$

Contrast this with the Thesis-compliant conditional probability:

$$P(D \Rightarrow B) = P(B|D) = P(B|DF)P(F|D) + P(B|D\bar{F})P(\bar{F}|D) \quad (13)$$

⁷The reader might want to pause for a moment to consider an alternative scenario where all the vases do in fact have a 7/10 probability of breaking if dropped. (Suppose, for instance, that there is a 3/10 chance of a soft landing which even the fragile vases would survive, and a 7/10 chance of landing in a gravel crusher which would break even the sturdy ones.) In this case, a vase's preciousness is independent of both whether it is dropped and whether it breaks; thus $P(P|DB) = P(P) = 2/10$, and so for the overall probability of the conditional we arrive at the following instead of (9):

$$2/10 \times P(D) + 2/10 \times P(\bar{D}) = 2/10 \quad (10)$$

⁸Since $P(DB) = P(DBF) + P(DB\bar{F}) = P(B|DF)P(DF) + P(B|D\bar{F})P(D\bar{F})$.

Suppose now that this locally interpreted conditional $D \Rightarrow_F B$ is itself the antecedent of a conditional $(D \Rightarrow_F B) \Rightarrow P$.⁹ Its probability is not equivalent to that of $(D \Rightarrow B) \Rightarrow P$, but can also be derived via a straightforward calculation. As a preliminary step, notice that the local interpretation obeys the probabilistic version of Conditional Excluded Middle (that is, the equivalence of $P(\overline{A \Rightarrow_F B})$ and $P(A \Rightarrow_F \overline{B})$):

$$\begin{aligned} & P(D \Rightarrow_F \overline{B}) + P(D \Rightarrow_F B) \\ &= (P(\overline{B}|DF) + P(B|DF))P(F) + (P(\overline{B}|D\overline{F}) + P(B|D\overline{F}))P(\overline{F}) \\ &= P(F) + P(\overline{F}) = 1 \end{aligned} \quad (14)$$

This ensures that the Fraction Lemma applies, that is, $\sum_n P(D \Rightarrow_F \overline{B})^n = 1/P(D \Rightarrow_F B)$. Thus:

$$\begin{aligned} & P((D \Rightarrow_F B) \Rightarrow P) \quad (15) \\ &= P\left(\bigcup_n \overline{D \Rightarrow_F B}^n \times \left(\frac{(DBP \times W^*) \cup \bigcup_n (\overline{D}FP \times \overline{D\overline{F}}^n \times DFB \times W^*)}{\bigcup_n (\overline{D}\overline{F}P \times \overline{D\overline{F}}^n \times D\overline{F}B \times W^*)}\right)\right) \\ &= \sum_n P(\overline{D \Rightarrow_F B})^n \cdot \left(\frac{P(DBP) + P(\overline{D}FP)\sum_n P(\overline{D\overline{F}}^n)P(DFB) + P(\overline{D}\overline{F}P)\sum_n P(\overline{D\overline{F}}^n)P(D\overline{F}B)}{P(\overline{D}\overline{F}P)\sum_n P(\overline{D\overline{F}}^n)P(D\overline{F}B)}\right) \\ &= \frac{P(DBP) + P(\overline{D}FP)P(B|DF) + P(\overline{D}\overline{F}P)P(B|D\overline{F})}{P(B|DF)P(F) + P(B|D\overline{F})P(\overline{F})} \quad (16) \end{aligned}$$

Now the case of the vase allows for further simplifications. Since $P(B|DF) = 1$ and $P(B|D\overline{F}) = 0$,

$$P((D \Rightarrow_F B) \Rightarrow P) = \frac{P(DBP) + P(\overline{D}FP)}{P(F)} \quad (17)$$

And since $P(DBP) = P(DFP)$ ¹⁰ we can simplify further:

$$P((D \Rightarrow_F B) \Rightarrow P) = \frac{P(DFP) + P(\overline{D}FP)}{P(F)} = \frac{P(FP)}{P(F)} = 2/7 \quad (18)$$

The end result is Bacon's desired probability of 2/7. Not only that, but the calculations took us through exactly the steps outlined in Bacon's informal paraphrase.

Notice also that while the assumption that fragility is not a matter of degree simplified the calculations, the applicability of the method did not depend on that assumption. This is significant, lest we be accused of merely substituting an

⁹Like [18], I do not assume that there is a local interpretation for conditionals with conditional antecedents, since I am not convinced that conditionals (or their denotations) are the relata of causal (in)dependence relations. Hence the unadorned second arrow in this and the following formulas.

¹⁰To see this, note that $P(DBP) = P(DFBP) + P(D\overline{F}BP)$; furthermore, $P(DFBP) = P(DF)P(B|DF)P(P|DFB)$; and since $P(B|DF) = 1$ and $P(P|DFB) = P(P|DF)$ due to conditional independence, this equals $P(DF)P(P|DF) = P(DFP)$. Lastly, $P(D\overline{F}BP) = 0$ because the sturdy vases don't break.

equivalent non-conditional sentence for the conditional antecedent (i.e., F for $D \Rightarrow B$ in this case). The claim that speakers perform such substitutions has at times been used in attempts to explain away the fact that conditionals can be embedded ([11] is a well-known example). But in our case it is merely a convenient coincidence that Bacon's description implied extreme conditional probabilities for the conditionals given the fragility of the vase. The approach does not depend on it.

Throughout this illustration I have used Bacon's "ur-conditional" $\Rightarrow$. Bacon also proposes a way to parametrize this operator to bodies of evidence E , obtaining a class of evidence-sensitive connectives $\Rightarrow_E$. This move would give him the wherewithal to fix the present problem, although doing so would require a reconceptualization of the role of E .

In Bacon's implementation, E is an accessibility relation on the set of possible worlds. At each world w , E_w is the set of worlds compatible with the evidence available to the agent at w . Bacon remarks that E should plausibly obey some introspection constraints, but stops short of endorsing a particular set thereof (p. 149-150). In the interpretation of conditionals, E enters as a parameter of the selection function: $f_E(A, w) = f(A \cap E_w, w)$. Thus in effect, at each world the conditional is interpreted as if its antecedent were conjoined with the evidence available at that world.

Technically, this idea of conjoining the antecedent at each world w with a certain proposition that is true at w is exactly what Kaufmann does to obtain his local interpretation. The difference is a conceptual one, concerning where this additional information comes from and what it represents. For Kaufmann, it embodies the values of relevant variables that are causally independent of the antecedent. The possible (combinations of) values of these variables partition the probability space, and an agent may give positive subjective probability to more than one cell in this partition. When this happens, the conditional has a local probability (the weighted sum of the conditional probabilities within each of the cells) in addition to its global probability (the conditional probability under the agent's information state). The two can come apart because a sum of ratios is not always equal to the ratio of the sums.¹¹

4 Deeply conditional antecedents

The last section showed that an Ordinal Semantic model would need to be modified to account for Bacon's own motivating example, but that the required fix involves a fairly straightforward importation of some of Kaufmann's ideas. This raises another question: is the additional formal complexity of Ordinal Semantics ever needed, or is Bernoulli Semantics all we need in models for the interpretation of conditionals?

Bacon himself is vague on this question. The closest thing to an empirical data point that he presents is the vase example, which as we saw can be considered an

¹¹ [22] adopts the basic idea and argues that the relevant partition is determined by salient questions in discourse rather than causal independencies. It may well be true that contextual factors can play this role at least in some cases. I take Khoo's argument as an overall endorsement of the approach; the empirical question of what determines the partition in particular cases is not conclusively resolved.

argument for a Kaufmann-style local interpretation, but not against the Bernoulli approach. Bacon claims that his account has an advantage on conditionals with even more deeply embedded antecedents, but he gives no details on the purported limitations of Bernoulli models. It is not clear whether he maintains that they fail to assign probabilities to such sentences, or that they assign the wrong probabilities to them. Nor does he say what the correct probabilities would be in these cases, or what probabilities they would receive under *his* account. None of the references he gives establishes any clear limitations of Bernoulli Semantics either.¹²

The literature has generally been unclear on this issue. Early on, [33] only established that Bernoulli models (he called them *Stalnaker Bernoulli models*) support the Thesis for sentences of the form $A \Rightarrow (B \Rightarrow C)$ and $(A \Rightarrow B) \Rightarrow C$, where A, B, C do not themselves contain conditionals. He added this caveat: “[A]lthough I do not know that the Thesis fails in them for more complicated conditionals I expect that it does” (p. 280). But this remained a mere conjecture until fairly recently, when [27] offered a concrete case in which a Bernoulli model indeed assigns a Thesis-violating probability to a sentence. The sentence in question is given in (19):

$$((\overline{B} \Rightarrow A) \Rightarrow \overline{B}) \Rightarrow A \quad (19)$$

As a piece of natural-language semantics, the formula seems contrived, even unnatural. Nor does [27] claim that any English sentence has (19) as its logical form. But that does not subtract from its value in establishing that Bernoulli models are reliably Thesis-conforming only up to a point.

I will only briefly outline the argument. At its center are three facts. The first is that, in a Bernoulli model, at any world sequence at which the antecedent of (19) – that is, the sentence $(\overline{B} \Rightarrow A) \Rightarrow \overline{B}$ – is false, (19) is true. The second fact is that the conditional probability of A given $(\overline{B} \Rightarrow A) \Rightarrow \overline{B}$ – that is, the Thesis-conforming probability of (19) – can in principle be strictly less than 1. The third is that in a Thesis-conforming Bernoulli model, if the probability of any conditional is less than 1, then it must be false at *some* set of sequences with positive probability at which its antecedent is false. These three facts are jointly incompatible.

The first observation is crucial because it highlights the source of the problem. The proof is tedious but straightforward. Suppose $(\overline{B} \Rightarrow A) \Rightarrow \overline{B}$ is false at w^* up to $w^*(n-1)$ for some n , then true at $w^*(n)$. This means that at $w^*(n-1)$, $\overline{B} \Rightarrow A$ is true and $\overline{B}$ false. Thus A is true at the first $\overline{B}$ -world after $w^*(n-1)$. The first $\overline{B}$ -world after $w^*(n-1)$ is either $w^*(n)$ or else the first $\overline{B}$ -world after $w^*(n)$. Either way, since $\overline{B} \Rightarrow A$ is true at $w^*(n-1)$, it is also true at $w^*(n)$. But then, since $(\overline{B} \Rightarrow A) \Rightarrow \overline{B}$ is true at $w^*(n)$, so is $\overline{B}$ – and thus, since $\overline{B} \Rightarrow A$ is true at $w^*(n)$, so is A . This completes the proof.

The proof also indicates why the example is a problem for Bernoulli Semantics. The crux is that $w^*(n-1)$ and $w^*(n)$ are immediately adjacent for the purposes of evaluating conditionals of any rank – necessarily so since there is no rank higher

¹²Bacon correctly cites [16] as ruling out conditional antecedents, but in the cited paper Jeffrey was not referring to Bernoulli models. [32] explicitly propose their version of Bernoulli Semantics for a language with unrestricted embeddings of conditionals.

than 1 in Bernoulli Semantics. Thus any conditional whose antecedent is false at $w^*(n-1)$ and true at $w^*(n)$, receives at $w^*(n-1)$ the truth value of its consequent at $w^*(n)$. For that reason, in the example, the truth of $\bar{B} \Rightarrow A$ at $w^*(n-1)$ forces the truth of A at $w^*(n)$, and thus ultimately the truth of (19) at w^* .

Put differently, the problem is an interaction between antecedents of different ranks. This does not happen in Ordinal Semantics: There, $\pi^\alpha(n-1)$ and $\pi^\alpha(n)$ are adjacent only at Rank α , but infinitely far apart at lower ranks. In the evaluation of $\bar{B} \Rightarrow A$, the relevant $\bar{B}$ -world following $\pi^\alpha(n-1)$ lies strictly between $\pi^\alpha(n-1)$ and $\pi^\alpha(n)$, and so there is no interaction between the truth of $\bar{B} \Rightarrow A$ at $\pi^\alpha(n-1)$ and the truth of A at $\pi^\alpha(n)$.

In conclusion, while Bernoulli models may be adequate up to a point, they can be driven to inconsistency in more complex applications. Ordinal models do not face this limitation.

5 Calculating probabilities

Having gone in some detail through arguments involving deeply embedded conditional antecedents, at this point one may wonder what probabilities are actually assigned to such sentences. Bacon's result (3) above means that the Thesis is obeyed at all ranks in Ordinal Semantics. But while the probabilities of conditionals (of arbitrary complexity) are determined by the probabilities in the underlying simple model and can be calculated from the latter, Bacon does not provide us with recipes for such calculations.

Deeply embedded conditionals like McCutcheon's (19) may rarely occur in everyday English, but it is not impossible to conceive of them: (20) is a case in point.

If it is not a suitable material if it breaks if subjected to stress, it won't be used. (20)

Intuitions about the probability of (20) may be hard to come by, but for what it is worth, its probability in an Ordinal Semantic model is as follows:

$$\begin{aligned} & P(((A \Rightarrow B) \Rightarrow C) \Rightarrow D) \\ &= \frac{P(ABCD) + P(B|A)P(\bar{A}CD)}{P(C|AB)P(A) + P(\bar{A}C)} + P(A\bar{B}D) + P(\bar{B}|A)P(\bar{A}D) \end{aligned} \quad (21)$$

This probability can be calculated via successive applications of the formulas in (22) through (24), derived by [18] for the case of non-conditional constituents. I use bold-faced letters in the display here to highlight the fact that they may in general contain further conditionals.

$$P(\mathbf{A} \Rightarrow \mathbf{B}) = P(\mathbf{B}|\mathbf{A}) = P(\mathbf{A} \cap \mathbf{B})/P(\mathbf{A}) \quad (22)$$

$$P((\mathbf{A} \Rightarrow \mathbf{B}) \cap (\mathbf{C} \Rightarrow \mathbf{D})) = \frac{P(\mathbf{ABCD}) + P(\mathbf{D}|\mathbf{C})P(\mathbf{AB}\bar{\mathbf{C}}) + P(\mathbf{B}|\mathbf{A})P(\mathbf{CD}\bar{\mathbf{A}})}{P(\mathbf{A} \cup \mathbf{C})} \quad (23)$$

$$P(\bar{\mathbf{A}} \Rightarrow \bar{\mathbf{B}}) = P(\mathbf{A} \Rightarrow \bar{\mathbf{B}}) \quad (24)$$

The formula in (21) instantiates the schema in (22), hence:

$$P(((A \Rightarrow B) \Rightarrow C) \Rightarrow D) = \frac{P(((A \Rightarrow B) \Rightarrow C) \cap D)}{P((A \Rightarrow B) \Rightarrow C)} \quad (25)$$

The numerator in (25) can be made to fit the schema for conjunctions of conditionals by inserting an arbitrary tautology ‘ $\top$ ’ as an antecedent for D . Then the conjunction rule (23) applies:¹³

$$\begin{aligned} & P(((A \Rightarrow B) \Rightarrow C) \cap (\top \Rightarrow D)) \\ &= \frac{\left[P((A \Rightarrow B) \cap C \top D) \right. \\ &\quad \left. + P(D|\top)P((A \Rightarrow B) \cap C \overline{\top}) + P(C|A \Rightarrow B)P(\top D \cap \overline{A \Rightarrow B}) \right]}{P((A \Rightarrow B) \cup \top)} \\ &= P((A \Rightarrow B) \cap C D) + P((A \Rightarrow B) \Rightarrow C)P(\overline{A \Rightarrow B} \cap D) \end{aligned} \quad (26)$$

The terms in (26) can all be simplified by further applications of the rules in (22) through (24):

$$\begin{aligned} P((A \Rightarrow B) \cap C D) &= P((A \Rightarrow B) \cap (\top \Rightarrow C D)) \\ &= P(ABCD) + P(B|A)P(\overline{A}CD) \end{aligned} \quad (27)$$

$$\begin{aligned} P((A \Rightarrow B) \Rightarrow C) &= P((A \Rightarrow B) \cap C)/P(A \Rightarrow B) \\ &= P((A \Rightarrow B) \cap (\top \Rightarrow C))/P(A \Rightarrow B) \\ &= [P(ABC) + P(B|A)P(\overline{A}C)]/P(B|A) \\ &= P(C|AB)P(A) + P(\overline{A}C) \end{aligned} \quad (28)$$

$$\begin{aligned} P(\overline{A \Rightarrow B} \cap D) &= P((A \Rightarrow \overline{B}) \cap (\top \Rightarrow D)) \\ &= P(A\overline{B}D) + P(\overline{B}|A)P(\overline{A}D) \end{aligned} \quad (29)$$

Substituting these back into (25) we derive (21):

$$\begin{aligned} & P(((A \Rightarrow B) \Rightarrow C) \Rightarrow D) \\ &= \frac{P((A \Rightarrow B) \cap C D) + P((A \Rightarrow B) \Rightarrow C)P(\overline{A \Rightarrow B} \cap D)}{P((A \Rightarrow B) \Rightarrow C)} \\ &= \frac{P((A \Rightarrow B) \cap C D)}{P((A \Rightarrow B) \Rightarrow C)} + P(\overline{A \Rightarrow B} \cap D) \\ &= \frac{P(ABCD) + P(B|A)P(\overline{A}CD)}{P(C|AB)P(A) + P(\overline{A}C)} + P(A\overline{B}D) + P(\overline{B}|A)P(\overline{A}D) \end{aligned} \quad (30)$$

¹³The simplification to the last line goes through since $P(\overline{\top}) = 0$, $P((A \Rightarrow B) \vee \top) = 1$, and $P(C|A \Rightarrow B) = P((A \Rightarrow B) \Rightarrow C)$.

6 Conclusions

Embedded conditionals are often taken to pose a challenge for the Thesis, but we know now that they do not. Ordinal Semantics is a powerful and very general framework for their analysis. Bernoulli Semantics is a simpler variant, but subject to breakdown with sentence that exceed a certain complexity. Both are adequate for the analysis of the linguistic examples that have thus far been discussed in the literature.

But just as the Thesis has never been uncontroversial even for first-order conditionals, the predictions of Ordinal or Bernoulli Semantics for complex conditionals are not always intuitively correct. Such discrepancies show that a Thesis-compliant model theory does not in itself amount to a probabilistic semantic analysis for conditionals; rather, it is merely a necessary ingredient of the latter, to be supplemented by an account of the other factors. Causal relations play a role, so does discourse structure (see the [Appendix](#) for an example). Mapping these interactions is the next challenge in the probabilistic semantics of conditionals.

Acknowledgments I am grateful to the editor and the reviewer for very helpful comments on an earlier version. All remaining errors and misrepresentations are my own. This work was supported in part by the National Science Foundation (#2116972, “Research on conditional and modal language”).

Appendix: Khoo on left-nested conditionals

An anonymous reviewer suggested that I address a recent criticism of the Bernoulli approach to left-nested conditionals of the form $(A \Rightarrow B) \Rightarrow C$. The criticism concerns cases in which none of A , B , C contain further embedded conditionals, thus the issue is not directly relevant to the comparison between Bernoulli Semantics and Ordinal Semantics since both agree on those conditionals. Nevertheless, I welcome this opportunity to discuss the criticism because it purports to identify an incorrect prediction of Bernoulli Semantics (hence presumably also of Ordinal Semantics). I will argue that the purported weakness is in fact an advantage.

The criticism is presented in [24] and [23]. I largely base my discussion on the latter. Khoo’s approach relies on sequences of worlds and bears some resemblances to Bernoulli semantics, but there are also important differences. For Khoo, the sequences represent Stalnakerian similarity orderings, rather than series of random picks. Thus a Khoo sequence contains each world exactly once, in contrast to Bernoulli sequences, in which worlds may occur repeatedly or not at all.

In Khoo’s framework, as in Bernoulli Semantics, first-order conditionals are interpreted by evaluating the consequent at the first world in the sequence at which the antecedent is true. However, Khoo introduces special interpretation rules for both right-nested and left-nested conditionals. The former need not concern us here.¹⁴ The latter presents a challenge for his overall framework. While Khoo does not assume that conditionals are propositional – in his models, like in Bernoulli models, they

¹⁴Khoo modifies the interpretation rule for right-nested conditionals to enforce the Import-Export Principle. See [18, 19] for relevant discussion.

can differ in truth value between sequences with the same base world – he does assume that their antecedents are propositional. That requirement is problematic for antecedents which themselves contain conditionals, as was first pointed out by [30]. I am not going to repeat that argument here (see also [10], Section 6.5). Bernoulli Semantics does not have the same problem because it does not require antecedents to be propositional. In Khoo’s framework, however, the upshot is that the Thesis cannot hold in general for left-nested conditionals.¹⁵ Khoo argues that this is as it should be, on the grounds that the Thesis-compliant conditional probability is not always intuitively correct.

This last claim deserves some discussion because it has some *prima facie* plausibility. But it is ultimately on the wrong track (or so I argue). Khoo produces two examples in which left-nested conditionals receive probabilities that are different from what is predicted under the Bernoulli approach. As in Bacon’s vase example discussed in this paper, at issue is again the question whether such perceived violations are counterexamples to the semantic account, or rather evidence of an interaction between the standard semantics and contextually given information. One of Khoo’s examples is exactly parallel to one discussed at length in Kaufmann ([18], Section 5.6.1) and susceptible to the same analysis. Here I focus on the other example.

Poker Paul. Paul is playing poker against Nancy. Nancy has a weak hand, but it is still possible Paul’s is weaker. It is also possible for Paul to win even if he has the weaker hand, but since Nancy is a good player, that is unlikely. Cheating is a way to increase your success of winning with a weaker hand, and Paul is not opposed to cheating, though it is unlikely he cheated if he had the better hand. But, if Paul won with the weaker hand, it is overwhelmingly likely that he cheated. Now consider:

If Paul won if he had the weaker hand, he cheated. (A1)

Khoo argues that (A1) is “very likely” in this situation. For concreteness he supplies the following numbers, along with the corresponding paraphrases.¹⁶

$\Pr(H) = .25$ (Paul likely does not have the weaker hand) (A2)

$\Pr(W|H) = .2$ (Paul likely did not win if he had the weaker hand)

$\Pr(C|\bar{H}) = .2$ (Paul likely did not cheat if he did not have the weaker hand)

$\Pr(C|HW) = .95$ (Paul likely cheated if he had the weaker hand and won)

¹⁵Much of Khoo’s discussion centers around the validity of the principle (CSO), which plays a role in Stalnaker’s result:

$$A \Rightarrow B, B \Rightarrow A, A \Rightarrow C \models B \Rightarrow C \quad (\text{CSO})$$

Khoo seeks to uphold (CSO) because he finds it intuitively valid. I agree with the intuition, but in a probabilistic setting I take the relevant notion to be *p*-validity: roughly speaking, an inference is *p*-valid if the conclusion cannot be unlikely while all the premises are likely [1, 3]. And (CSO) is *p*-valid (see [6], for a recent proof).

¹⁶These sentences are cited verbatim from Khoo, including the disorienting oscillation between Past and Present tense. Given his wording, it is not clear whether the game is still ongoing or already completed. I will largely stay with Past tense in my own subsequent discussion.

Based on these numbers, the conditional probability of C given $H \Rightarrow W$ is given in (A4). We already saw that this is also the probability that the conditional (A1) receives in a Bernoulli model with these probabilities.

$$\begin{aligned}\Pr(C|H \Rightarrow W) &= \Pr(C|HW)P(H) + \Pr(C|\overline{H}) \\ &= \Pr(C|HW)P(H) + \Pr(\overline{H})\Pr(C|\overline{H}) \\ &= .95 \times .25 + .75 \times .2 = .3875\end{aligned}\quad (\text{A3})$$

In Khoo's judgment, this is too low for the conditional: he finds that (A1) is "very likely" in the given scenario. On the other hand, it is less clear (to me) what Khoo thinks of the conditional probability of C , given $H \Rightarrow W$. He presents the example as evidence against the Thesis, thus he apparently accepts that the conditional probability is .3875. However, his justification for the claim that the conditional is very likely sounds much like a paraphrase of his reasoning about C after (hypothetically) learning $H \Rightarrow W$: "Paul's winning if he had the weaker hand most likely had to do with his cheating (since he was likely not good enough to beat Nancy with a weaker hand)." If this is Khoo's reasoning in evaluating the conditional but not in calculating the conditional probability, I would like to know more about the difference between them. Absent that, I am not sure whether he thinks that both $\Pr(C|H \Rightarrow W)$ and $\Pr((H \Rightarrow W) \Rightarrow C)$ should be high, or only the latter.

However that may be, my contention is that both of these probabilities are in fact .3875 – higher than the prior probability that Paul cheated, but not much higher.¹⁷ Thus in this case, learning that Paul won if he had the weaker hand raises the probability that he cheated from less than .245 to .3875. I suspect that Khoo's intuition that the probability ought to be higher is based on reasoning that involves a step not licensed by the scenario as given. Suppose the conditional antecedent is true. I concur with Khoo's intuition that "Paul's winning if he had the weaker hand most likely had to do with his cheating", but we cannot conclude that he likely cheated. All we can conclude is that he likely cheated *if he had the weaker hand*. For the scenario explicitly stipulates that whether Paul cheated depended on his hand; thus the supposition that $H \Rightarrow W$ is true does not affect the (low) conditional probability that he cheated in the (likely) event that he had the stronger hand.

¹⁷While Khoo does not give us $\Pr(C)$, some bounds are implied. First, note that $\Pr(H)$ and $\Pr(HWC)$ are fixed at .25 and .0475, respectively:

$$\Pr(HWC) = \Pr(H)\Pr(W|H)\Pr(C|HW) = .25 \times .2 \times .96 = .0475 \quad (\text{i})$$

From this it follows that $\Pr(W|HC)$ and $\Pr(C|H)$ are inversely proportional:

$$.0475 = .25 \times \Pr(C|H) \times \Pr(W|HC) \quad (\text{ii})$$

In other words, the more likely it is that Paul cheated if he had the weaker hand, the less likely it is that that cheating was successful. Now, Khoo does say that cheating with the weaker hand raises Paul's chances of winning, but he does not say how high or by how much. Assuming that cheating with the weaker hand makes him more likely to win than to lose (i.e., $\Pr(W|HC) > .5$), we must conclude that Paul is rather unlikely to have cheated if he had the weaker hand ($\Pr(C|H) < .38$), and the prior probability that he cheated is less than .245:

$$\Pr(C) = \Pr(H)\Pr(C|H) + \Pr(\overline{H})\Pr(C|\overline{H}) < .25 \times .38 + .75 \times .2 = .245 \quad (\text{iii})$$

Thus while only Khoo knows for sure what is in Khoo's mind, my best guess is that he implicitly interprets the conditional along the lines of (A4).

If Paul won if he had the weaker hand, he cheated [if he had the weaker hand].
(A4)

The two conditional constituents of (A4) share the same antecedent. For this reason, the formula for calculating the probability of the sentence can be simplified as in (A6). The result is indeed a much higher probability.

$$\begin{aligned} & \Pr((H \Rightarrow W) \Rightarrow (H \Rightarrow C)) \\ &= \frac{\Pr(HWC) + \Pr(C|H)\Pr(HW\bar{H}) + \Pr(W|H)\Pr(\bar{H}HC)}{\Pr(H \cup H)\Pr(W|H)} \\ &= \Pr(HWC)/\Pr(HW) = \Pr(C|HW) = .95 \end{aligned} \quad (A5)$$

From a linguistic perspective, it is not implausible that the conditional should be prone to reinterpretation in this way. It has long been observed that conditional antecedents, aside from their truth-conditional role within their sentences, tend to set up *topics* which participate in the structuring of the surrounding discourse context ([7, 13, 28], i.a.). In the case of (A4), the overarching discourse topic would be a certain property of propositions – specifically, the property of *being true if Paul had the weaker hand*. Within this context, the sentence states, in effect, that if *Paul won* has this property, then so does *Paul cheated*. The fact that the scenario stipulates that whether Paul cheated depended on his hand may facilitate this interpretation.

To be sure, embedded conditionals have not been discussed in the linguistic literature on topichood and discourse structure, and the above argument must remain speculative at this point. It certainly warrants further investigation.¹⁸

Be that as it may, it is instructive to compare this hypothesis to Khoo's own way around the problem with (A1). Khoo pursues two goals: to treat the conditional antecedent of $(A \Rightarrow B) \Rightarrow C$ as propositional, and to derive a high probability for (A1) in the scenario given above. His proposed solution is a radical kind of contextualism, according to which the antecedent denotes different propositions depending on the world sequence (not just the world) at which it is interpreted. Specifically, a conditional $A \Rightarrow B$ which occurs in a conditional antecedent is interpreted as either the material conditional $A \supset B$ or the conjunction AB :¹⁹

Khoo's pointwise interpretation

$$(A \Rightarrow^K B) \Rightarrow C \text{ is interpreted at } \sigma \text{ as : } \begin{cases} C & \text{if } A \Rightarrow B \text{ is true at } \sigma \\ AB \Rightarrow C & \text{otherwise} \end{cases} \quad (A6)$$

¹⁸It is worth noting in this connection that a link between conditionals of the form $(A \Rightarrow B) \Rightarrow C$ and the conditional probability $\Pr(C|AB)$ has been suggested before by [9].

¹⁹Khoo's definition says that the conditional is interpreted as $(A \supset B) \Rightarrow C$ at sequences at which $A \Rightarrow B$ is true. But that collapses into C by strong centering and since $A \Rightarrow B$ entails $A \supset B$. Khoo does not seem to take this into account, as a result the probabilities he derives are slightly off.

The expectation of these values is as in (A7); details are given in a footnote.²⁰

$$\begin{aligned} \Pr((A \Rightarrow^K B) \Rightarrow C) \\ = \Pr(ABC) + \Pr(\bar{A}C)\Pr(B|A) + \Pr(C|AB)\Pr(\bar{B}|A) \end{aligned} \quad (\text{A7})$$

In the given scenario, this comes out to the probability in (A8).

$$\begin{aligned} \Pr((H \Rightarrow W) \Rightarrow C) \\ = \Pr(HWC) + \Pr(C\bar{H})\Pr(W|H) + \Pr(C|HW)\Pr(\bar{W}|H) \\ = .0475 + .15 \times .2 + .95 \times .8 = .8375 \end{aligned} \quad (\text{A8})$$

Incidentally, (A7) is equivalent to (A9), which highlights the similarity to the $\Pr(C|AB)$ I suggested above:^{21,22}

$$\Pr(C|AB)\Pr(A) + [\Pr(C|\bar{A})\Pr(B|A) + \Pr(C|AB)\Pr(\bar{B}|A)]\Pr(\bar{A}) \quad (\text{A9})$$

In sum, while I think Khoo has a point regarding the intuitive probability of the sentence, this judgment is not a counterargument to Bernoulli Semantics. On the contrary, I maintain that Bernoulli Semantics is correct and the judgment is the result of an interaction with pragmatic factors. Khoo's ad-hoc semantic redefinition does not do justice to either of those components.

²⁰First, the sequences at which $A \Rightarrow B$ is true fall into two disjoint subsets, AB and $\bar{A} (A \Rightarrow B)$. Thus:

$$\begin{aligned} \Pr(((A \Rightarrow^K B) \Rightarrow C) \wedge (A \Rightarrow B)) \\ = \Pr(ABC) + \Pr(\bar{A}C)[\Pr(AB) + \Pr(\bar{A})\Pr(AB) + \Pr(\bar{A})^2\Pr(AB) + \dots] \\ = \Pr(ABC) + \Pr(\bar{A}C)\Pr(B|A) \end{aligned} \quad (\text{i})$$

Second, at sequences at which $A \Rightarrow B$ is false, so is AB , thus the value of $AB \Rightarrow C$ is that of C at the first AB -world.

$$\begin{aligned} \Pr(((A \Rightarrow^K B) \Rightarrow C) \wedge (\bar{A} \Rightarrow \bar{B})) \\ = \Pr(C|AB)[\Pr(A\bar{B}) + \Pr(\bar{A})\Pr(A\bar{B}) + \Pr(\bar{A})^2\Pr(A\bar{B}) + \dots] \\ = \Pr(C|AB)\Pr(\bar{B}|A) \end{aligned} \quad (\text{ii})$$

The probability in (A7) is the sum of (i) and (ii).

²¹Proof:

$$\begin{aligned} (\text{A9}) &= \Pr(C|AB)\Pr(A) + \Pr(\bar{A}C)\Pr(B|A) + \Pr(C|AB)\Pr(\bar{B}|A)\Pr(\bar{A}) \\ &= \Pr(C|AB)\Pr(A) + \Pr(\bar{A}C)\Pr(B|A) \\ &\quad + \Pr(C|AB)\Pr(\bar{B}|A) - \Pr(C|AB)\Pr(\bar{B}|A)\Pr(A) \text{ since } \Pr(\bar{A}) = 1 - \Pr(A) \\ &= \Pr(C|AB)\Pr(A) + \Pr(\bar{A}C)\Pr(B|A) + \Pr(C|AB)\Pr(\bar{B}|A) \\ &\quad - \Pr(C|AB)\Pr(A) + \Pr(C|AB)\Pr(B|A)\Pr(A) \text{ since } \Pr(\bar{B}|A) = 1 - \Pr(B|A) \\ &= \Pr(\bar{A}C)\Pr(B|A) + \Pr(C|AB)\Pr(\bar{B}|A) + \Pr(ABC) = (\text{A7}) \end{aligned}$$

²²Notice that Khoo's probability will generally be close to $\Pr(C|AB)$:

$$\Pr(C|AB) = \Pr(C|AB)\Pr(A) + [\Pr(C|AB)\Pr(B|A) + \Pr(C|AB)\Pr(\bar{B}|A)]\Pr(\bar{A})$$

References

1. Adams, E. (1965). The logic of conditionals. *Inquiry*, 8, 166–197.
2. Adams, E. (1975). *The Logic of Conditionals*. Dordrecht: Reidel.
3. Adams, E. (1998). *A Primer of Probability Logic*. Stanford: CSLI Publications.
4. Bacon, A. (2015). Stalnaker's thesis in context. *The Review of Symbolic Logic*, 8(1), 131–163.
5. Bennett, J. (2003). *A Philosophical Guide to Conditionals*. London: Oxford University Press.
6. Berto, F., & Özgün, A. (2021). Indicative conditionals: Probabilities and relevance. *Philosophical Studies*, 178, 3697–3730.
7. Ebert, C., Ebert, C., & Hinterwimmer, S. (2014). A unified analysis of conditionals as topics. *Linguistics and Philosophy*, 37, 353–408.
8. Edgington, D. (1986). Do conditionals have truth conditions? *Critica*, 18(52), 3–30.
9. Edgington, D. (1991). The mystery of the missing matter of fact. In *Proceedings of the Aristotelian Society*, pp. 185–209, *The Aristotelian Society*.
10. Edgington, D. (1995). On conditionals. *Mind*, 104(414), 235–329.
11. Gibbard, A. (1981). Two recent theories of conditionals. pp. 211–247. In W. L. Harper, R. Stalnaker, & G. Pearce (Eds.) *Ifs: Conditionals, Belief, Decision, Chance, and Time*. Dordrecht: Reidel.
12. Grice, H. P. (1989). *Studies in the Way of Words*. Cambridge: Harvard University Press.
13. Haiman, J. (1978, September). Conditionals are topics. *Language*, 54(3), 564–589.
14. Häjek, A., & Hall, N. (1994). The hypothesis of the conditional construal of conditional probability. pp. 75–110. In E. Eells, & B. Skyrms (Eds.) *Probabilities and Conditionals: Belief Revision and Rational Decision*. Cambridge: Cambridge University Press.
15. Jackson, F. (1979). On assertion and indicative conditionals. *Philosophical Review*, 88, 565–589.
16. Jeffrey, R. C. (1991). *Matter-of-fact conditionals*. In *The Symposia Read at the Joint Session of the Aristotelian Society and the Mind Association at the University of Durham. Supplementary* (Vol. 65, pp. 161–183). London: The Aristotelian Society.
17. Kaufmann, S. (2005). Conditional predictions: A probabilistic account. *Linguistics and Philosophy*, 28(2), 181–231.
18. Kaufmann, S. (2009). Conditionals right and left: Probabilities for the whole family. *Journal of Philosophical Logic*, 38, 1–53.
19. Kaufmann, S. (2015). Conditionals, conditional probabilities, and conditionalization. In H. C. Schmitz, & H. Zeevat (Eds.) *Bayesian Natural Language Semantics and Pragmatics* (pp. 71–94). Springer.
20. Kaufmann, S. (2017a). Probabilistic semantics and pragmatics for the language of uncertainty. In M. B. Ferraro, P. Giordani, B. Vantaggi, M. Gagolewski, M. Á. Gil, P. Grzegorzewski, & O. Hryniewicz (Eds.) *Soft Methods for Data Science*, (Vol. 456 of *Advances in Intelligent Systems and Computing*, pp. 285–291). Springer.
21. Kaufmann, S. (2017b). Towards a probabilistic analysis for conditionals and unconditionals. In M. Otake, S. Kurahashi, K. Satoh, & D. Bekki (Eds.) *New Frontiers in Artificial Intelligence*, (Vol. 10091 of *Lecture Notes in Computer Science*, pp. 3–14). Springer.
22. Khoo, J. (2016). Probabilities of conditionals in context. *Linguistics and Philosophy*, 39(1), 1–43.
23. Khoo, J. (2021). *The meaning of If_i. Ms.* Cambridge: MIT. Version of October 7, 2021.
24. Khoo, J., & Santorio, P. (2018). Lecture notes: Probability of conditionals in modal semantics. Last accessed on June 3, 2022. <https://www.justinkhoo.org/s/KS-NASSLLI2018.pdf>.
25. Lewis, D. (1976). Probabilities of conditionals and conditional probabilities. *Philosophical Review*, 85, 297–315.
26. Lewis, D. (1986). Probabilities of conditionals and conditional probabilities II. *Philosophical Review*, 95, 581–589.
27. McCutcheon, R. G. (2019). Adams thesis and the local interpretation of conditionals. <http://philsci-archive.pitt.edu/16321/>. Last accessed on June 3, 2022.
28. Schiffrin, D. (1992). Conditionals as topics in discourse. *Linguistics*, 30, 165–197.
29. Stalnaker, R. (1968). A theory of conditionals. In Rescher, N. (Ed.), *Studies in Logical Theory, American Philosophical Quarterly, Monograph: 2*, (pp. 98–112). Oxford: Blackwell.
30. Stalnaker, R. (1976). Letter to van Fraassen. pp. 302–306. In W. L. Harper, & C. A. Hooker (Eds.) *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science, Vol 1 of The University of Western Ontario Series in Philosophy of Science*. D. Reidel.

31. Stalnaker, R. (1981). A defense of conditional excluded middle. pp. 87–104. In W. L. Harper, R. Stalnaker, & G. Pearce (Eds.) *Ifs: Conditionals, Belief, Decision, Chance, and Time*. Dordrecht: Reidel.
32. Stalnaker, R., & Jeffrey, R. (1994). Conditionals as random variables. pp. 31–46. In E. Eells, & B. Skyrms (Eds.) *Probabilities and Conditionals: Belief Revision and Rational Decision*. Cambridge: Cambridge University Press.
33. van Fraassen, B. C. (1976). Probabilities of conditionals. pp. 261–308. In W. L. Harper, & C. A. Hooker (Eds.) *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science, Vol 1 of The University of Western Ontario Series in Philosophy of Science*. D. Reidel.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.