Pragmatic intrusion in probability judgment: The case of conditionals

Daniel Lassiter (dan.lassiter@ed.ac.uk)
School of Philosophy, Psychology, and Language Sciences
University of Edinburgh

Chunan Li (c.li-112@sms.ed.ac.uk)
School of Philosophy, Psychology, and Language Sciences
University of Edinburgh

Abstract

Recent research has provided experimental support for a new "Inferentialist" theory of conditionals, challenging the Equation $P(\text{If } A, C) = P(C \mid A)$ and theories that support it. The key evidence comes from probability judgments involving conditionals whose antecedent and consequent are relevant vs. irrelevant to each other. Expanding on recent experimental work, we argue that Inferentialism has difficulty explaining the data. However, theories that support The Equation theory are well-placed to account for the results once we recognize an independent phenomenon of pragmatic intrusion on probability judgment—in this case, participants' tendency to assign lower probability to conditionals that are pragmatically incoherent.

Keywords: Conditionals, probability judgment, pragmatics, discourse coherence

Conditionals and their probabilities

Theories of conditionals drawn from linguistics and philosophy play a crucial role in psychological research by providing testable hypotheses and theoretical frames for interpreting behavioral data. The detailed survey of Evans and Over (2004) argues that the suppositional theory of conditionals provides the best available account of participants' behavior in a wide range of experiments. Inspired by Ramsey (1929), this theory holds that we evaluate a conditional "If A, C" by temporarily assuming that the antecedent A is true, and considering the status of the consequent C in this light.

Ramsey's idea can be implemented formally in a variety of ways (e.g. Kratzer, 1991; Stalnaker, 1978). In current psychological research, the best-supported implementation is the trivalent or "defective truth-table" account proposed independently by de Finetti (1936) and Wason (1966) (see also Over & Cruz, 2021; Politzer, Over, & Baratgin, 2010, a.m.o.). The trivalent semantics has also had a revival in philosophical and linguistic research (Égré, Rossi, & Sprenger, 2021; Huitink, 2008). In this theory, a conditional with a false antecedent is undefined; when the antecedent is true, the conditional has the truth-value of its consequent.

Wason (1966) proposed a trivalent theory to account for participants' tendency to treat false-antecedent conditionals as "irrelevant", rather than true or false. For de Finetti, a primary motivation was to explain the intuitive equality between probabilities of conditionals and conditional probabilities:

$$P(\text{If } A, C) = P(C \mid A) \tag{1}$$

"The Equation" (1), as Edgington (1995) dubs it, has much conceptual and experimental support (e.g., Adams, 1975; Douven & Verbrugge, 2010; Evans & Over, 2004; Oberauer & Wilhelm, 2003). While Lewis (1976) famously proved that the Equation cannot hold in a bivalent semantics, the proof does not apply in a trivalent setting. Eq. (1) holds in a trivalent theory that defines $P(S)$ as the probability that S is true, divided by the probability that S is defined (Cantwell, 2006; Lassiter, 2020; Milne, 1997). It is supported by the suppositional theory and also by a recent version of Mental Models theory (López-Astorga, Ragni, & Johnson-Laird, 2022).

Trouble for the Equation

Recent results cast doubt on whether the Equation is correct, and so on theories that are designed to predict it. Skovgaard-Olsen, Singmann, and Klauer (2016) and Douven, Elqayam, and Mirabile (2022) showed experimentally that probability judgments of "If A, C" do not always track judgments of $P(C \mid A)$. We focus here on the latter work, since our experiment builds directly upon it. Stimuli illustrating the key comparison are given in (2-3), where the (b)-judgments are intended as a measure of $P(C \mid A)$.

(2) a. What is the probability of this sentence: "If the US cuts its oil consumption, global warming will slow down."

b. Suppose that the US cuts its oil consumption. In your opinion, how likely is it that global warming will slow down?

Douven et al. found that the two judgments types were well-correlated for conditionals like (2). However, in stimuli like (3) where A and C are irrelevant to each other, judgments diverged: probability ratings (3a) were generally lower.

(3) a. What is the probability of this sentence: "If robots become shy, Ukraine will be allowed to join the EU."

b. Suppose that robots become shy. In your opinion, how likely is it that Ukraine will be allowed to join the EU?

Prima facie, this result would appear to disconfirm Eq. 1.

Douven et al. also investigated a third experimental measure that was designed to track strength of inferential connection between A and C: "Suppose A. How strongly do you agree that it then follows that C?" They found that this measure is substantially better at predicting $P(\text{If } A, C)$. This result

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

provides support for **Inferentialism**, an alternative semantics that is much-discussed in recent work (e.g. Douven, 2008; Douven, Elqayam, Singmann, & van Wijnbergen-Huitink, 2018; Krzyżanowska, Collins, & Hahn, 2017; Krzyżanowska & Douven, 2018; Skovgaard-Olsen, 2016). This theory holds that the conventional meaning of a conditional consists in the requirement that the consequent can be inferred from the antecedent. If so, our willingness to accept a conditional should track the strength of the inferential connection between its antecedent and consequent.

Examples like (3) provide a key test of Inferentialism against the Equation (1). Suppose that someone thinks that Ukraine will very likely be allowed to join the EU, and that this is totally unconnected to the (rather strange) issue about robot shyness. When C and A are independent, $P(C \mid A) = P(C)$. So, the suppositional theory predicts that this person should rate both (3a) and (3b) highly. In contrast, Inferentialism predicts—correctly—that this person may rate (3b) highly while giving a low rating to (3a). On this theory, this happens when they think it is unlikely that the consequent can be inferred from the antecedent. This is essentially what drives the main result of Douven et al. (2022).

Another way to see the problem is to consider a formal consequence of the Equation (1) that will be crucial later on—what we call the "Intermediacy" condition.

$$\begin{aligned} P(C) &= P(A)P(C \mid A) + P(\neg A)P(C \mid \neg A) \\ &= P(A)P(\text{“If } A, C\text{”}) + P(\neg A)P(\text{“If not } A, C\text{”}). \end{aligned} \tag{4}$$

Since $P(A) + P(\neg A) = 1$, $P(C)$ is a convex combination of P(“If A, C”) and P(“If not A, C”). So the Equation implies

> **Intermediacy**: The probability of the consequent C is bounded by P(“If A, C”) and P(“If not A, C”): it cannot be greater than both, or less than both.

The experiment of Douven et al. (2022) does not demonstrate violations of Intermediacy, because they did not test matched pairs of positive (If A, C) and negative (If not A, C) conditionals, or gather ratings of the consequent C. However, their results lead us to expect that we should find such violations at least in examples like (3a). In particular, Inferentialism predicts that both P(If A, C) and P(If not-A, C) should be lower than $P(C)$ in such conditionals. The experiment reported below confirms this prediction of Inferentialism. Positive/negative/consequent triples based on (3a) do violate Intermediacy, falsifying the Equation for examples like (3a).

Two kinds of irrelevance

While this reasoning appears compelling, it fails to recognize a crucial pragmatic distinction between two kinds of irrelevance. The first is the sort of generalized irrelevance that we see in (3a), where there is no conceivable connection between A and C. Over (2023) and Cruz and Over (2024) dub these "Walrus conditionals", in reference to Carroll's (1872) Walrus: "The time has come ... to talk of many things: of shoes—and ships—and sealing-wax—and cabbages—and kings ...".

The conditional in (3) does not merely show independence between A and C. As Lassiter (2023) points out, the key examples used to motivate Inferentialism are **incoherent** in a technical sense drawn from discourse pragmatics: there is no way to establish a connection between the two clauses that could support a coherent discourse, in the sense of Asher and Lascarides (2003); Hobbs (1979); Kehler (2002); Knott and Dale (1994). (See Cruz, Over, Oaksford, and Baratgin 2016 for a related point.) In this respect, there is nothing special in (3) about the fact that the sentence is a conditional. Connecting the same clauses with another device (*and*, *even though*, simple juxtaposition) would generate a similar bizarreness. Lassiter argues that conditionals like (2a) are often coherent by virtue of instantiating the coherence relations **Result** or **Explanation**, which involves a causal or inferential connection. For instance, a conjunctive variant of (2a) strongly implies that the event in the second clause occurred *because of* the event in the first, even though this is not stated explicitly.

(4) The U.S. cut its oil consumption, and global warming slowed down.

A different type of irrelevance involves examples in which A and C could be connected, but the actual context is one in which they are not. For instance, in (5)—a stimulus from the experiment below—it is easy to imagine how the clauses could be connected by a **Result** relation. However, the context sentence denies the potential connection, rendering the clauses probabilistically and inferentially unrelated.

(5) [Context: Alicia's umbrella is broken.] If Alicia's umbrella was in her bag yesterday, she did not get wet in the unexpected rainshower.

The example is pragmatically coherent, but the contextual information plausibly renders the antecedent and consequent independent: most likely, Alicia got wet no matter where her umbrella was. We will call examples like (5) "Contextual Independence" conditionals. From the perspective of discourse coherence, these conditionals should show very different pragmatic behavior from incoherent Walrus conditionals.

Alternative account: Pragmatic intrusion

The fact that Walrus conditionals are pragmatically incoherent does not yet explain why they received low probability ratings in the experiment of Douven et al. (2022). Discourse coherence is a pragmatic concept, and normative probability judgment involves the semantic interpretation of a sentence. To account for the existing results within (e.g.) the suppositional theory, we need an additional linking hypothesis:

> **Linking theory**. Probability judgments may take into account the entire message that an utterance conveys, not just literal meaning. As a result, pragmatic infelicity—e.g., incoherence—tends to lower probability ratings.

To account for the results of Douven et al. (2022), we must also assume that "suppose" judgments are not affected.

This linking theory is supported by previous experimental results. Cremers, Križ, and Chemla (2017) found that participants often took pragmatic factors into account in making probability judgments. Fugard, Pfeifer, and Mayerhofer (2011) found a large difference between the felicitous (6a) and the nearly equivalent, but pragmatically infelicitous, (6b).

(6) a. If the card shows a 2, it shows an even number.
b. If the card shows a 2, it shows a 2 or a 4.

While most participants assigned probability 1 to (6a), most judged (6b) to have probability 0—even though it is plainly true. Fugard et al. (2011) attribute this effect to the fact that (6b) violates pragmatic norms (Quantity and Manner: see Grice 1975). Relatedly, many experiments show pragmatic intrusion in truth-value judgments (e.g. Bott & Noveck, 2004; Degen & Tanenhaus, 2015; De Neys & Schaeken, 2007).

Distinguishing Predictions

So far, Inferentialism has an advantage, straightforwardly predicting the existing data: Walrus conditionals have low probability because they are false due to lack of an inferential connection. Theories that rely on the Equation have access to an adequate, but more complex, account in terms of pragmatic intrusion. The theories do, however, come apart in their predictions about an issue that has not yet been tested: the probabilities of Contextual Independence conditionals. Such conditionals can be pragmatically coherent, even though there is no inferential connection between the clauses. The basic predictions of the competing theories appear to be:

- **Inferentialism**: Contextual Independence conditionals should behave like Walrus conditionals, since both fail to show an inferential connection between antecedent and consequent. As a result, $P(\text{If } A, C)$ and $P(\text{If } \neg A, C)$ should be systematically lower than $P(C)$, violating Intermediacy.
- **Equation-cum-pragmatic-intrusion**: Contextual Independence conditionals do not have any pragmatic defect, and so they should not behave like pragmatically incoherent Walrus conditionals. Instead, judgments of $P(\text{If } A, C)$ and $P(\text{If } \neg A, C)$ should be approximately equal to $P(C)$. Contextual Independence conditionals should not violate Intermediacy, but Walrus conditionals should.

Our experiment tests these predictions with judgments involving these three types of conditionals across the three question types used in the experiment of Douven et al. (2022).

Experiment

Participants, materials and procedure

We recruited 100 participants from Prolific.com, restricted to monolingual native speakers of English in the US or UK. The experiment was administered on a locally hosted website built using jsPysch (de Leeuw, Gilbert, & Luchterhandt, 2023). Completion took 6 minutes on average. Particpants were compensated £1.20 for their time (mean payment £12/hour).

Our experiment built heavily on the design of Douven et al. (2022), but added a crucial manipulation. In addition to 10 Walrus conditionals borrowed from that experiment, we designed 10 Dependence conditionals (their "Positive-link" conditionals). For each Dependence conditional we designed an optional "defeater"—a piece of background information that should sever the expected connection between antecedent and consequent, rendering them independent. This manipulation allowed us to compare responses to conditionals across trials differing only in whether a defeater was present.

Participants were randomly assigned to one of three conditions: **Probability**, **Suppose**, or **InfStrength**. Examples of trials in the three conditions are shown in Fig. 1. The same items occurred in all three conditions, adapted slightly for the format of the questions. For example, the second item in the middle panel of Fig. 1 appeared in the Suppose condition as "Suppose the sun has gone down. In your opinion, how likely is it that it is dark in Paul's bedroom?" In the InfStrength condition, the same item appeared as "Suppose that the sun has gone down. How strongly do you agree that it then follows that it is dark in Paul's bedroom?"

For the Probability task, we adapted the method from exp.2 of Douven et al. (2022). On each trial, participants rated the probability of three sentences on a slider with endpoints labeled with "Certainly false" and "Certainly true". The first was the bare consequent C. The second was the Positive conditional ("If A, C"), and the third was the Negative conditional ("If not A, C"). After selecting a response for each sentence, participants were permitted to proceed to the next trial.

Following Douven et al. (2022), the Suppose task was intended as an operationalization of conditional probability judgment. The format of these trials was identical to the Probability task except that the wording of the prompts differed: Participants rated the likelihood of the consequent, and then rated the same sentence under positive and negative suppositions (see Fig. 1). Ratings were collected on a slider with endpoints labeled (following Douven et al. (2022)) with "Very unlikely" and "Very likely". For comparability with results of Douven et al. (2022), we retained the qualifier "In your opinion" in each of the "Suppose" questions.

In the InfStrength task participants judged strength of inferential connection between A and C. We again used question wording and slider labels taken directly from Douven et al. 2022. Because inferential strength is a relation between two sentences, it was not possible to include a matched question about C, as we had in the other two conditions.

Each participant saw a total of 15 trials of two main types: 5 Walrus and 10 non-Walrus, in random order. Each participant's 5 Walrus items were picked randomly from among 10 items borrowed from Douven et al. (2022), with small modifications to improve the English and avoid interference from events that had occurred since that experiment was conducted. We also designed 10 non-Walrus conditionals similar to (5), loosely based on stimuli from Byrne, Espino, and Santamaria 1999. These conditionals were chosen as ones where

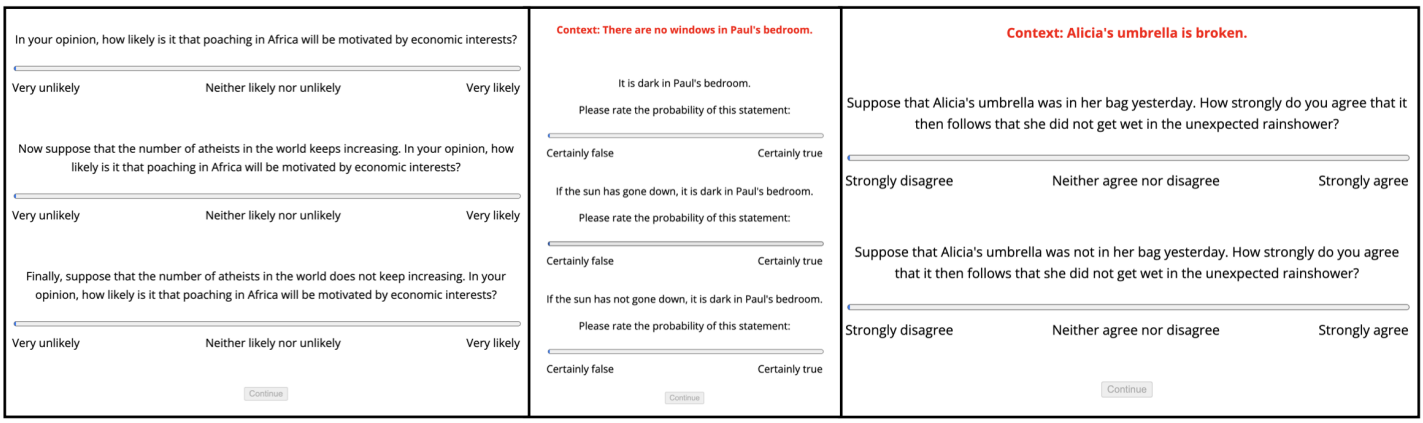

Figure 1: Example trials from the **Suppose**, **Probability**, and **InfStrength** conditions.

antecedent and consequent would normally be inferentially related, but it was possible to deny this connection with a carefully chosen defeater. In the 10 non-Walrus trials, the contextual defeater was shown with probability .5, independently per trial. When a defeater appeared, it was placed at the top and highlighted in bold red font to encourage participants to attend to it (see Fig. 1, middle and right). Participants thus saw 5 Walrus trials and, on average, 5 dependence and 5 Contextual Independence trials, with some random variation in the precise mixture of the latter two types.

Results

Our experiment collected 4,035 slider ratings from 1,500 trials: (465 InfStrength, 600 Probability, 435 Suppose, with variation due to random assignment). Due to an implementation error, we were forced to discard data from 58 trials for two non-Walrus items (6 and 10).

We first compare responses for Dependence and Contextual Independence items. If the contextual defeater manipulation was successful, differences between positive, negative, and consequent ratings for Dependence items should be neutralized in their Contextual Independence counterparts, which include a defeater (red line, the middle and right panels of Fig. 1). Fig. 2 plots the mean absolute difference between positive and negative ratings for each item type and condition. The inclusion of a contextual defeater rendered positive and negative judgments very similar in all conditions—nearly as small as those of Walrus conditionals.

We now proceed to test the predictions of the competing theories (see "Distinguishing Predictions" above). Our analysis will center around the Intermediacy condition because it is a useful test for whether participants' judgments are consistent with the predictions of the Equation. In other words, Intermediacy is a coherence requirement on Equation-conforming probability judgments (cf. Evans, Thompson, & Over, 2015). Theories that support the Equation predict that it should be satisfied, modulo experimental noise; theories that deny the Equation predict that it should have systematic

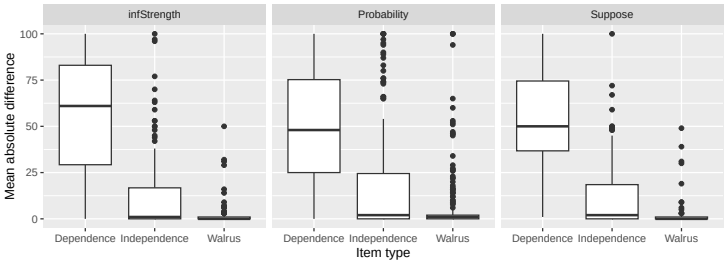

Figure 2: Mean absolute difference between positive and negative prompts, by condition.

counter-examples in some cases.

Fig. 3 plots the overall response patterns and means in all nine conditions. For Dependence conditionals, mean judgments satisfy Intermediacy in both Probability and Suppose conditions. In Walrus items, Positive and Negative conditionals have substantially lower mean Probability ratings than their consequents, in violation of Intermediacy. Specifically, mean probability judgments for positive and negative conditionals were much lower than for their consequents (positive = 26.7, negative = 22.1, consequent = 51.9). In contrast, mean Suppose judgments were virtually identical for positive, negative, and consequent for Walrus items (mean positive = 50.4, mean negative = 50.3, mean consequent = 53.2). These broad patterns are consistent with the predictions of both theories.

The key issue is whether Contextual Independence items pattern with Dependence items or with Walrus items. As Fig. 3 shows, mean responses to Contextual Independence items conform to Intermediacy in both Probability and Suppose judgments, as predicted by the pragmatic intrusion theory. The mean response patterns do not point to reduced Probability ratings for Contextual Independence items, as Inferentialism would predict.

These results are not yet probative, since Intermediacy is a prediction about the behavior of individual participants re-

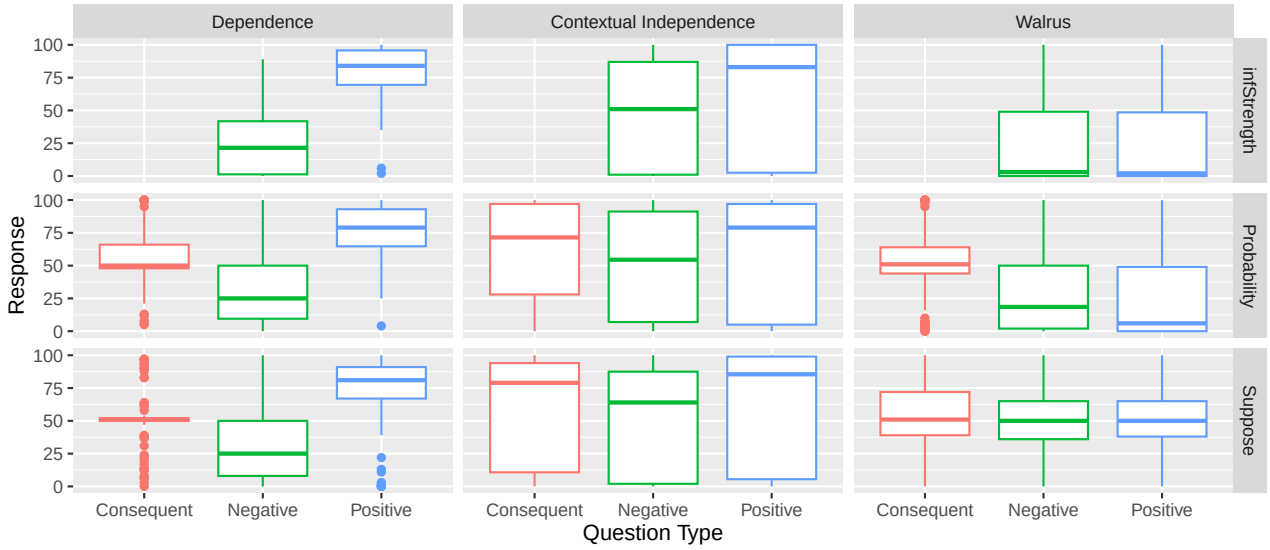

Figure 3: Mean responses for all conditions and question types.

sponding to the content of specific items. We thus proceed to an item-by-item analysis with attention to responses in individual trials. As expected, for all 10 Dependence items, mean responses satisfied Intermediacy in both Probability and Suppose conditions, and participants' judgments did so in 68% of individual trials in the Probability condition and in 73% of trials in the Suppose condition.

For all ten Walrus items, mean responses in the Probability condition violated Intermediacy, with much lower ratings for Positive and Negative conditionals than for their consequents. By contrast, mean Suppose ratings for Walrus items were virtually identical for Consequent, Positive, and Negative questions. Probability judgments for Walrus items satisfied Intermediacy in just 21% of individual trials, while Suppose judgments did so in 48% of trials.

For the Contextual Independence conditionals, item-by-item analysis showed no consistent pattern of reduced probability for positive and negative conditionals relative to their consequents. Mean responses satisfied Intermediacy in six of 10 items, and were approximately consistent in three more. The divergences in the latter cases were consistent with noisy responding: they all occurred for items where the three judgments were numerically very close, and did not exceed 5% of the scale. Only item 1 suggested a clear violation of Intermediacy (mean consequent = 35.6; mean negative = 22, mean positive = 11.8). This was also the only item in the experiment whose mean Suppose ratings violated Intermediacy, suggesting a confound (see footnote).[1] Probability ratings for Contextual Independence conditionals satisfied Intermediacy in 52% of individual trials, and Suppose judgments did so in 60% of trials. This pattern is in stark constrast to Walrus conditionals, where Intermediacy failed in mean responses for all ten items and in a much higher proportion of individual trials.

To confirm these findings, we performed nested model comparisons using mixed-effects logistic regression models, predicting the probability that a given trial satisfied Intermediacy from judgment type (Probability vs. Suppose) and item type (Dependence, Contextual Independence, or Walrus). This analysis excluded InfStrength trials, where we did not have consequent data. The first comparison, including Dependence trials, had by-item and by-participant random intercepts (the models with interaction terms did not converge with random slopes). The model comparison revealed a highly significant interaction between judgment type and item type ($\chi^2(2) = 11.328$, $p < .005$). The interaction model also had significant main effects of item type for both Contextual Independence items ($\beta = -0.70$, SE = .249, $p < .005$) and Walrus items ($\beta = -2.26$, SE = .388, $p < 10^{-13}$), and a significant interaction between condition and item type for Walrus items only ($\beta = 1.07$, SE = .37, $p < .005$).

A second comparison tested the crucial prediction that Walrus and Contextual Independence items differed in their propensity to conform to Intermediacy, excluding data from the Dependence condition. Again, we performed a nested model comparison, predicting the probability that a given trial satisfied Intermediacy from judgment type (Probability vs. Suppose) and item type (Contextual Independence

[1] Item 1 was likely confounded as follows. It had the contextual defeater "Someone has removed the batteries from Bill's remote control.", and the consequent question read "How likely is it that Bill's TV will turn on?". Some participants may have assigned a relatively high probability to the consequent (mean = 35.6) because they considered the possibility of turning on the TV using an different method, e.g., a button on the TV itself—an interpretation which would not be pragmatically available for the corresponding positive and negative conditionals, which led "If Bill presses/does not press the button on his remote control, ...". This account would also explain why this item was unique in violating Intermediacy in the Suppose condition—an unexpected effect for Inferentialism and the suppositional theory alike.

or Walrus). Again we were able to use only by-participant random slopes. Here too, the model comparison showed a highly significant interaction between judgment type and item type ($\chi^2(1) = 8.8469$, $p < .005$). The full interaction model showed a highly significant main effect of item type ($\beta = -1.61$, SE=0.25, $p < 10^{-9}$) and a highly significant interaction between item type and judgment type ($\beta = -1.1$, SE=0.37, $p < .005$). The finding of an interaction confirms our earlier inspection of the results. Judgment type (Probability vs. Suppose) makes a larger difference in how likely Walrus conditionals are to conform to Intermediacy than it does for Contextual Independence conditionals.

Discussion

The key prediction tested in our experiment involved probability judgments for our Contextual Independence conditionals, and how they compared to Dependence and Walrus conditionals. Inferentialism predicts that Contextual Independence conditionals should resemble Walrus conditionals in having reduced probability for positive and negative conditionals relative to the probability of their consequents, because of a failure of inferential connection. In contrast, the pragmatic-intrusion theory and the Equation (1) together predict that positive and negative Contextual Independence conditionals should not show pragmatic intrusion because they do not violate discourse coherence. As a result, this theory predicts that the probabilities of Contextual Independence conditionals should be approximately equal to the probabilities of their consequents. The theory also predicts that Contextual Independence conditionals should conform to Intermediacy, a formal consequence of the Equation $P(\text{“If } A, C\text{”}) = P(C \mid A)$. In our analysis, we used conformity to this condition as a way to test whether participants' pattern of judgments were consistent with the Equation.

The results supported the pragmatic intrusion theory. Predictions shared by the two theories were confirmed. However, Contextual Independence conditionals, both positive and negative, were generally rated as having probability approximately equal to the probability of their consequents—as pragmatic intrusion predicts—and not as having lower probability than their consequents, as Inferentialism predicts. This result was in stark contrast to the results for Walrus conditionals, where $P(\text{If } A, C)$ and $P(\text{If not } A, C)$ were systematically lower than $P(C)$ across all ten items tested (see fig. 3).

These results support the pragmatic intrusion theory's hypothesis that the reduced probability of Walrus conditionals can be attributed to their lack of discourse coherence. It also disconfirms the Inferentialist hypothesis that this effect can be attributed to a lack of inferential connection: no similar effect was observed in Contextual Independence conditionals, even though they also lacked an inferential connection.

One possible response for an advocate of Inferentialism is to deny that the theory makes the predictions about Contextual Independence conditionals described in section “Distinguishing Predictions” above. Specifically, advocates of Inferentialism have distinguished “normal” conditionals—to which the theory is meant to apply—from “special” conditionals, which are semantically different and outside the scope of the theory (e.g., Douven 2008; Skovgaard-Olsen 2016). However, this way of eliminating the problematic data from consideration is not theoretically well-motivated. In particular, there is no independent evidence that the semantic interpretation of these conditionals is different from those of Dependence and Walrus conditionals. Nor is there any known mechanism from natural language semantics or pragmatics by which the addition of a piece of background information (e.g., “Alicia's umbrella is broken” in (5)) could modify the semantic interpretation of the word “if” in a subsequent sentence. We therefore believe that Inferentialists cannot avoid the problematic empirical predictions described above without adopting other methodologically and linguistically implausible positions.

Conclusion

Previous experimental results suggest that Gricean pragmatics can influence probability judgment. This paper provides, to our knowledge, the first experimental result showing that coherence relations (Hobbs, 1979; Kehler, 2002) also influence probability judgment. To show this, it was necessary to demonstrate that the reduced probabilities of Walrus conditionals are in fact due to a pragmatic violation (failure of discourse coherence) and not to their conventional semantics. Our experimental results supported the pragmatic-intrusion theory and the Equation by showing that conditionals that are pragmatically coherent, but show no inferential connection between antecedent and consequent, are not associated with lowered probability judgments. This is in contrast with pragmatically incoherent Walrus conditionals, whose probabilities are systematically judged to be lower than the corresponding conditional probabilities.

The suppositional theory is popular in psychological work, as well as linguistics and philosophy. This theory is strongly associated with (1), the Equation between probabilities of conditionals and conditional probabilities. As a result, the apparent failure of the Equation in experiments by Douven et al. (2022) and Skovgaard-Olsen et al. (2016) threatens the suppositional theory, along with other theories that support the Equation (e.g., the Mental Models account of López-Astorga et al. (2022)). However, our results indicate that the failure is merely apparent, and can be attributed to the fact that the methods used to measure probabilities of conditionals in these experiments are also influenced by pragmatic factors. Properly interpreted, empirical data around probability judgments do support Equation (1) and theories that enforce it.[2]

[2]We are grateful to David Over, Simone Sebben, and three anonymous CogSci reviewers for extremely helpful feedback. All data and analysis scripts are available at the OSF archive linked here. This research was supported by a grant from the BA/Leverhulme Small Research Grants Scheme.

References

Adams, E. W. (1975). *The logic of conditionals: An application of probability to deductive logic*. Springer.

Asher, N., & Lascarides, A. (2003). *Logics of conversation*. Cambridge University Press.

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of memory and language*, *51*(3), 437–457.

Byrne, R. M., Espino, O., & Santamaria, C. (1999). Counterexamples and the suppression of inferences. *Journal of Memory and Language*, *40*(3), 347–373.

Cantwell, J. (2006). The laws of non-bivalent probability. *Logic and Logical Philosophy*, *15*(2), 163–171.

Carroll, L. (1872). *Through the looking glass and what alice found there*. Macmillan and Co.

Cremers, A., Križ, M., & Chemla, E. (2017). Probability judgments of gappy sentences. In *Linguistic and psycholinguistic approaches on implicatures and presuppositions* (pp. 111–150). Springer.

Cruz, N., & Over, D. E. (2024). From de finetti's three values to conditional probabilities in the psychology of reasoning. In P. E. . L. Rossi (Ed.), *Handbook of trivalent logics.* MIT Press.

Cruz, N., Over, D. E., Oaksford, M., & Baratgin, J. (2016). Centering and the meaning of conditionals. In *Proceedings of the thirty-seventh annual conference of the cognitive science society.*

de Finetti, B. (1936). La logique de la probabilité. In *Actes du congrès international de philosophie scientifique* (Vol. 4, pp. 1–9).

Degen, J., & Tanenhaus, M. K. (2015). Processing scalar implicature: A constraint-based approach. *Cognitive science*, *39*(4), 667–710.

de Leeuw, J. R., Gilbert, R. A., & Luchterhandt, B. (2023). jspsych: Enabling an open-source collaborative ecosystem of behavioral experiments. *Journal of Open Source Software*, *8*(85), 5351.

De Neys, W., & Schaeken, W. (2007). When people are more logical under cognitive load: Dual task impact on scalar implicature. *Experimental psychology*, *54*(2), 128–133.

Douven, I. (2008). The evidential support theory of conditionals. *Synthese*, *164*(1), 19–44.

Douven, I., Elqayam, S., & Mirabile, P. (2022). Inference strength predicts the probability of conditionals better than conditional probability does. *Journal of Memory and Language*, *123*, 104302.

Douven, I., Elqayam, S., Singmann, H., & van Wijnbergen-Huitink, J. (2018). Conditionals and inferential connections: A hypothetical inferential theory. *Cognitive Psychology*, *101*, 50–81.

Douven, I., & Verbrugge, S. (2010). The Adams family. *Cognition*, *117*(3), 302–318.

Edgington, D. (1995). On conditionals. *Mind*, *104*(414), 235-329.

Égré, P., Rossi, L., & Sprenger, J. (2021). De Finettian logics of indicative conditionals part I: Trivalent semantics and validity. *Journal of Philosophical Logic*, *50*(2), 187–213.

Evans, J. S. B. T., & Over, D. E. (2004). *If*. Oxford University Press.

Evans, J. S. B. T., Thompson, V. A., & Over, D. E. (2015). Uncertain deduction and conditional reasoning. *Frontiers in Psychology*, *6*, 129830.

Fugard, A. J., Pfeifer, N., & Mayerhofer, B. (2011). Probabilistic theories of reasoning need pragmatics too: Modulating relevance in uncertain conditionals. *Journal of Pragmatics*, *43*(7), 2034–2042.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics 9: Pragmatics* (pp. 41–58). Academic Press.

Hobbs, J. R. (1979). Coherence and coreference. *Cognitive science*, *3*(1), 67–90.

Huitink, J. (2008). *Modals, Conditionals and Compositionality*. Unpublished doctoral dissertation, Radboud University Nijmegen.

Kehler, A. (2002). *Coherence, reference, and the theory of grammar*. CSLI Publications.

Knott, A., & Dale, R. (1994). Using linguistic phenomena to motivate a set of coherence relations. *Discourse processes*, *18*(1), 35–62.

Kratzer, A. (1991). Modality. In A. von Stechow & D. Wunderlich (Eds.), *Semantik: Ein internationales Handbuch der zeitgenössischen Forschung* (pp. 639–650). Walter de Gruyter.

Krzyżanowska, K., Collins, P. J., & Hahn, U. (2017). Between a conditional's antecedent and its consequent: Discourse coherence vs. probabilistic relevance. *Cognition*, *164*, 199–205.

Krzyżanowska, K., & Douven, I. (2018). Missing-link conditionals: pragmatically infelicitous or semantically defective? *Intercultural Pragmatics*, *15*(2), 191–211.

Lassiter, D. (2020). What we can learn from how trivalent conditionals avoid triviality. *Inquiry*, *63*(9-10), 1087–1114.

Lassiter, D. (2023). Decomposing relevance in conditionals. *Mind & Language*, *38*(3), 644–668.

Lewis, D. (1976). Probabilities of conditionals and conditional probabilities. *Philosophical Review*, *85*(3), 297–315. doi: 10.2307/2184045

López-Astorga, M., Ragni, M., & Johnson-Laird, P. N. (2022). The probability of conditionals: A review. *Psychonomic Bulletin & Review*, *29*(1), 1–20.

Milne, P. (1997). Bruno de Finetti and the logic of conditional events. *The British Journal for the Philosophy of Science*, *48*(2), 195–232.

Oberauer, K., & Wilhelm, O. (2003). The meaning (s) of conditionals: Conditional probabilities, mental models, and personal utilities. *Journal of Experimental*

Psychology: Learning, Memory, and Cognition, *29*(4), 680.

Over, D. E. (2023). Independence and rationality. *The Reasoner*, *17*(6), 47-48.

Over, D. E., & Cruz, N. (2021). The suppositional theory of conditionals and rationality. In M. Knauff & W. Spohn (Eds.), *Handbook of rationality* (pp. 395–403).

Politzer, G., Over, D. E., & Baratgin, J. (2010). Betting on conditionals. *Thinking & Reasoning*, *16*(3), 172–197.

Ramsey, F. P. (1929). General propositions and causality. In *The foundations of mathematics and other logical essays* (p. 237-255). Kegan Paul.

Skovgaard-Olsen, N. (2016). Motivating the relevance approach to conditionals. *Mind & Language*, *31*(5), 555–579.

Skovgaard-Olsen, N., Singmann, H., & Klauer, K. C. (2016). The relevance effect and conditionals. *Cognition*, *150*, 26–36.

Stalnaker, R. (1978). Assertion. In P. Cole (Ed.), *Syntax and semantics 9: Pragmatics.* Academic Press.

Wason, P. C. (1966). Reasoning. In B. Foss (Ed.), *New horizons in psychology* (pp. 135–151). Penguin.