Inference strength predicts the probability of conditionals better than conditional probability does

Igor Douven, Shira Elqayam, Patricia Mirabile

HAL Id: hal-03508592
https://hal.science/hal-03508592

Submitted on 3 Jan 2022

Inference Strength Predicts the Probability of Conditionals Better than Conditional Probability Does

3 authors, including:

Igor Douven
French National Centre for Scientific Research
205 PUBLICATIONS   **2,896** CITATIONS

SEE PROFILE

Shira Elqayam
De Montfort University
47 PUBLICATIONS   **730** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Realism View project

Explanatory Reasoning View project

Inference Strength Predicts the Probability of Conditionals Better than Conditional Probability Does

Igor Douven[*]
IHPST / CNRS / Panthéon–Sorbonne University, France

Shira Elqayam
School of Applied Social Sciences, De Montfort University, United Kingdom

Patricia Mirabile
ILLC, University of Amsterdam, The Netherlands

Abstract

According to the philosophical theory of inferentialism and its psychological counterpart, Hypothetical Inferential Theory (HIT), the meaning of an indicative conditional centrally involves the strength of the inferential connection between its antecedent and its consequent. This paper states, for the first time, the implications of HIT for the probabilities of conditionals. We report two experiments comparing these implications with those of the suppositional account of conditionals, according to which the probability of a conditional equals the corresponding conditional probability. A total of 358 participants were presented with everyday conditionals across three different tasks: judging the probability of the conditionals; judging the corresponding conditional probabilities; and judging the strength of the inference from antecedent to consequent. In both experiments, we found inference strength to be a much stronger predictor of the probability of conditionals than conditional probability, thus supporting HIT.

Keywords: conditional probability; conditionals; Hypothetical Inferential Theory; inference; inferentialism; probability of conditionals; suppositional account.

1   Introduction

Conditionals are sentences of the form "If φ, [then] ψ," with φ called the "antecedent" and ψ, the "consequent." Conditionals have long been central to the study of human thinking and reasoning. But a comprehensive and empirically adequate theory of conditionals is still missing, and to this day there is little consensus on the meaning of conditionals or on the terms of their use in everyday conversation.

This paper focuses on *indicative* conditionals, such as "If you strike the match, it will light up," which are conditionals whose antecedent is in the indicative mood.[1] There is an ongoing controversy among researchers about the semantics of these statements: What do they mean? What are their truth conditions? Do they have truth conditions at all? Our aim is to provide further support for inferentialism, a recently proposed semantics (Douven, 2016a; Krzyżanowska,

---

[*]The three authors contributed equally to the paper and are listed alphabetically.
[1]Henceforth, we simply use "conditionals" to refer to indicative conditionals.

2015; Krzyżanowska, Wenmackers, & Douven, 2014), and for an even more recent psychological theory of conditionals, Hypothetical Inferential Theory or HIT (Douven, Elqayam, Singmann, & van Wijnbergen-Huitink, 2018, 2020), which builds on inferentialism. According to inferentialism, we judge whether a conditional is true or false by testing the inferential connection from its antecedent to its consequent. What HIT adds is the psychological mechanisms of representational and control processes: that the default relevant mental representation is that in which there is an inferential connection; and that this inference need only be strong enough in the sense suggested by Simon (1982).

1.1   Theoretical background

The suppositional account.   In philosophy, one of the most popular accounts of conditionals is that of the probability conditional (Adams, 1975, 1998; Edgington, 2014). Central to this account is what has been dubbed *the Equation*, according to which the probability of a conditional, $p(\text{If } \varphi, \psi)$, equals the corresponding conditional probability, $p(\psi \mid \varphi)$.[2] In psychology, some authors have adopted the philosophical version (Oaksford & Chater, 2003), while others (Evans, 2020; Evans & Over, 2004) have developed it further by linking it to the simulation heuristic in the study of judgment and decision-making (Kahneman & Tversky, 1982) and to the (presumably) uniquely-developed human ability to think hypothetically about possibilities removed in space and time. This has resulted in what is now commonly known as "the suppositional account." While philosophers advocating the probability conditional have done so mostly on the basis of its alleged pre-theoretical plausibility (see, e.g., van Fraassen, 1976), psychologists have reported various studies supporting the descriptive adequacy of the Equation; see Douven and Verbrugge (2010, 2013), Evans, Handley, and Over (2003), Fugard, Pfeifer, Mayerhofer, and Kleiter (2011), Gauffroy and Barrouillet (2009), Hadjichristidis et al. (2001), Oaksford and Chater (2003, 2007), Oberauer, Weidenfeld, and Fischer (2007), Over, Douven, and Verbrugge (2013), Over and Evans (2003), and Politzer et al. (2010).

Recently, however, Skovgaard-Olsen, Singmann, and Klauer (2016) have shown the Equation to break down for conditionals whose antecedent is not probabilistically positively relevant to its consequent. That should not have come as a complete surprise, if only because the condition of positive relevance is unlikely to hold for so-called missing link conditionals, which are conditionals lacking an apparent connection between their constituent parts (Douven, 2017a; Krzyżanowska & Douven, 2018); for example, "If you strike a match, the Liberal Democrats will win the next general election in the UK."[3] Indeed, it has often been observed that conditionals, in order to be acceptable or assertable, appear to require some sort of connection between

---

[2]Proponents of the suppositional theory do not all agree on the same semantics for conditionals. For instance, Adams (1975, 1998) and Edgington (2014) hold that conditionals do not express propositions and so are not the kind of things that can be true or false. On the other hand, Politzer, Over, and Baratgin (2010) propose a three-valued semantics for conditionals, on which a conditional is true, false, or indeterminate, and show how the Equation can be derived from that semantics. Both accounts face problems, but we will not go into these here and refer interested readers to Douven (2016a, Ch. 2) for arguments against the former account and to Douven (2016b) for arguments against the latter.

[3]Some critics of inferentialism appear to believe that the absence of an inferential connection can be *defined* in terms of probabilistic irrelevance; see, for instance, Over and Cruz (in press). That is a mistake, however. While we cannot think of a missing link conditional whose component parts are not probabilistically independent of each other, that does not mean that whenever a conditional's component parts are probabilistically independent of each other, that conditional is a missing link conditional. Consider a coin with unknown bias; the bias could be anything. Then the probability that the coin will land heads given that it is fair equals 0.5, which is also the unconditional probability that the coin will land heads. Nevertheless, "If the coin is fair, it will land heads" is *not* a missing link conditional.

their antecedent and consequent (e.g., Berto & Özgün, in press; Crupi & Iacona, in press; Douven, 2008; Krzyżanowska, 2019; Rott, 1986; Skovgaard-Olsen, 2016; Spohn, 2015; van Rooij & Schulz, 2019). The suppositional account seems unable to do justice to that intuition, given that the Equation is to hold for any φ and ψ, however unrelated they may be. Advocates of the account could try to fill this gap by an appeal to pragmatics, perhaps arguing that the seeming requirement of a connection arises from general conversational principles of the sort discussed by Grice (1989). But there are both theoretical and empirical reasons to doubt that such attempts will be successful; see Douven (2008, 2016a) for the former and Douven and Krzyżanowska (2019), Krzyżanowska, Collins, and Hahn (2017, 2021), Krzyżanowska and Douven (2018), and Rostworowski, Pietrulewicz, and Będkowski (in press) for the latter.

Inferentialism. This has led a number of philosophers to revisit an old idea about conditionals that had never been brought fully to fruition. The old idea is simply to make the requirement of an antecedent–consequent connection front and center of the semantics of conditionals: such a connection must be in place for the conditional to be *true*. The idea goes back to the Greek philosophers, in particular Chrysippus (Kneale & Kneale, 1962), and was later, in various forms, advocated by such eminent philosophers as Mackie (1973), Mill (1843), and Ramsey (1929/1990); in psychology, Braine (1978) and Braine and O'Brien (1991) have been notable defenders of the idea. What these authors' proposals have in common is that they postulate as a truth condition for a conditional the existence of an inferential relationship between that conditional's antecedent and consequent: the consequent must *follow* from the antecedent for the conditional to be true.

That this approach has never come fully to fruition, as said, is because most, or perhaps even all, of the aforementioned authors intended the requisite inferential connection to be understood as that of *entailment*, or *deductive* consequence. And if we require a conditional's consequent to follow deductively from its antecedent for that conditional to be true, then there are just too many conditionals that we intuitively regard as true that, in that approach, emerge as false. Too often, conditionals appear to express only a *defeasible* inferential connection (see, e.g., Delgrande, 1998). To cite an example from Douven et al. (2018, p. 52), it is not hard to imagine circumstances under which we would regard "If Betty misses her bus, she will be late for the movies" as true, even if those circumstances do not completely rule out the possibility of some coincidence through which, in the event that she missed her bus, Betty might still make it to the cinema in time.

Krzyżanowska et al. (2014) took this observation as a starting point for a new incarnation of the aforementioned approach, which they dubbed "inferentialism" (see also Douven, 2016a; Krzyżanowska, 2015; Krzyżanowska et al., 2021; Krzyżanowska, Wenmackers, & Douven, 2013; Mirabile & Douven, 2020; Oaksford & Chater, 2010, 2013, 2014, 2017, 2020; van Rooij & Schulz, 2019). In that proposal, for the conditional to be true, the consequent must still follow from the antecedent, but "follow" is no longer understood as entailment. Instead, the proposal is that the consequent must follow from the antecedent in a looser sense, where this looser sense may appeal to forms of non-deductive inference (including abductive, analogical, and inductive inference). More exactly, on this proposal there must be a compelling argument from the conditional's antecedent—in conjunction with background knowledge—to the conditional's consequent, where the antecedent should be a non-redundant premise in the argument. Importantly, for an argument to be compelling it need not be *conclusive*. In other words, the argument need not consist only of, and in fact need not involve *any*, deductive steps, but may involve inductive, abductive, and other non-deductive inferential steps. Arguably, what laypeople, with little or no

_____

(To forestall further misunderstanding, that does not mean it is *true*. There is an inferential connection between antecedent and consequent alright, but it is too weak to afford a compelling argument from the former to the latter.)

3

formal background in logic or mathematics, have in mind when they say that one thing follows from another is much closer to the sense in which Krzyżanowska and colleagues use this notion in their proposal than to the notion of entailment as formalized in classical logic.[4]

Because in this paper predictions from an approach based on inferentialism will be contrasted with ones based on a purely probabilistic approach, it is worth stressing that inferential strength, as understood by inferentialists, is not definable in probabilistic, or even logico-mathematical, terms. This is most easily seen, and most widely agreed upon, for abductive and analogical inference. In abductive inference, we reason to the best explanation, where explanation quality is typically understood in terms of the so-called theoretical virtues, including simplicity, mathematical elegance, scope, and coherence with background knowledge (Douven, 2017b, 2021a, in press). Few (if any) philosophers nowadays hold that all these notions, or even any of them, can be completely captured in formal terms. For the notion of explanation itself, this may even be more obvious, at least if we agree with modern authors that this notion is to be explicated in terms of understanding (de Regt, 2017) or articulated awareness (Woody, 2004). As for analogical inference, this is generally taken to exploit similarity relations (Carnap, 1980; Paris & Vencovská, 2017), and while there are various proposals around for formalizing such relations (e.g., Carnap, 1980; Decock & Douven, 2011; Gärdenfors, 2000; Douven, Elqayam, Gärdenfors, & Mirabile, in press), none of them suggest an obvious reduction of similarity to probability. Inductive inference, finally, exploits frequency information (Kyburg & Teng, 2001). Whereas this may raise the hope that induction can be defined purely probabilistically, arguments given in Nelkin (2000) and Douven and Williamson (2006) show that that hope would be vain.

Hypothetical Inferential Theory. Just as the probability conditional served as a key input for the suppositional account, inferentialism was one of the main building blocks of HIT (Douven et al., 2018, 2020), which tops up the philosophical apparatus with several psychological postulates, based on the dual process approach of hypothetical thinking theory (Evans, 2006, 2007a). In the sense suggested by Marr (1982), inferentialism provides the computational level of analysis for HIT, whereas hypothetical thinking theory provides the algorithmic level of analysis (i.e., processing and representations). Two principles are derived at this algorithmic level. At the representational level, the *principle of relevant inference* is the idea that the relevant mental representation of conditionals is by default the one in which there is an inferential relation between antecedent and consequent. At the processing level, the *principle of bounded inference* means this link need only be strong enough, in the sense of being subjectively supported, and not maximally strong. Thus, reasoners only invest processing effort to a limited degree, and satisfice on arguments that are just compelling enough. In short, HIT postulates that the key to understanding indicative condi-

---

[4]It is worth mentioning that, as presented in Krzyżanowska et al. (2014), inferentialism was meant to apply to standard indicative conditionals, and explicitly not to what authors have called "nonconditionals" (e.g., Geis & Lycan, 1993)—such as speech act conditionals ("If you're hungry, there are cookies on the table") and non-interference conditionals ("If hell freezes over, Alice will not marry Bob")—nor to subjunctive conditionals and concessives ("even if" conditionals, which are sometimes expressed without "even"; see Douven & Verbrugge, 2012). Thus criticisms of the position that accuse inferentialism of being unable to account for such nonconditionals (e.g., Mellor & Bradley, in press; Over & Cruz, in press) are misguided. This is not to say that it would be impossible to extend inferentialism to cover various other classes of conditionals. For instance, Douven (2016a, p. 38 f) points out that it would be rather straightforward to extend inferentialism to subjunctive conditionals. And a reasonable first stab at an inferentialist account of concessives is to define "[Even] if φ, ψ" to be true precisely if there is a compelling argument for ψ from background premises alone and also from those premises revised (in the sense of Alchourrón, Gärdenfors, & Makinson, 1985) with φ (i.e., given one's current background knowledge, there is a compelling argument from φ to ψ, but φ would serve as a redundant premise in that argument). See Douven, Elqayam, and Krzyżanowska (in press) for more on this.

4

tionals is a relevance-driven, satisficing-bounded inferential connection between antecedent and consequent.

Previous empirical work (Douven et al., 2018, 2020) sought to support HIT by presenting participants with two tasks. Participants were shown a soritical series of fourteen color patches gradually varying in color from clearly blue to clearly green, with adjacent patches very similar in color (see Fig. 5.1 in Douven et al., 2018). They were then asked to evaluate the truth value of conditionals pertaining to that series, such as "If patch number 3 is blue, so is patch number 6." The second task presented the same materials in argument form (with the antecedent as premise and the consequent as conclusion), for instance, "Suppose patch number 3 is blue; does it then follow that so is patch number 6?" Participants were instructed to evaluate the strength of inference from premise to conclusion as well as their metacognitive confidence in their own evaluation, both on a 7-point Likert scale. Both inference strength and metacognitive confidence turned out to be reliable and strong predictors of whether the corresponding conditionals would be evaluated as true, as was predicted by HIT.

To illustrate the relation between inferentialism and HIT, consider the curious case of belief bias in conditionals. Belief bias (Evans, Barston, & Pollard, 1983) is a well-documented interaction between inference strength (for example, logical validity) and believability of the conclusion. For example, given that all roses are birds, and all birds are insects, we should validly infer that all roses are insects. People still shy away from such inference because the conclusion, that all roses are insects, is unbelievable. Dual process theories explain this bias as a product of heuristic processing. Naturally, if conditionals are micro-inferences, we should expect belief bias in conditionals as well, so belief in the consequent (which is, by this analogy, the conclusion of the inference) should be a factor in truth evaluation of conditionals. The findings in Douven et al. (2018) strictly conformed to this predicted pattern. To make this prediction, we need both inferentialism as a philosophical theory and HIT as a psychological theory: the latter is essential for understanding that inference also entails bias.

The gap we aim to address in this paper concerns the implications of HIT for the probabilities of conditionals. Douven et al. (2018, 2020) had relegated the task of working out those implications to future work. The task is important in itself, but it is also important because complementing HIT with an account of the probabilities of conditionals will allow for a more direct comparison between HIT and the suppositional account than could be undertaken so far. A secondary aim of this paper is to extend HIT beyond its previous empirical paradigm, which drew on fairly abstract conditionals, and was limited to truth-table evaluations.

Hypothetical Inferential Theory and the probabilities of conditionals. Until now, HIT has not been extended to explaining how people estimate the probability of a conditional. This is an important lacuna which we now aim to fill in. In the past decades, the psychology of reasoning has seen the rise of a theoretical approach known as the *New Paradigm* (Elqayam, 2017; Elqayam & Over, 2013). This paradigm builds on a broadly Bayesian framework, which recognizes that we seldom treat any statement as purely true or false, but typically have confidence in statements to varying degrees (for variations within this approach, see Elqayam & Evans, 2013). This is in contrast to the more traditional paradigm, which saw classical logic as providing the norms of rational reasoning, hence focusing on categorical—and binary—truth and falsity. With the advent of the New Paradigm, researchers have focused as much or even more on probability assignments to conditionals than on categorical truth ratings.

Moreover, the implications of HIT for the probabilities of conditionals need to be addressed in view of the previously mentioned finding that the Equation appears to break down for missing

link conditionals. We submit that HIT has the potential to illuminate this negative finding. The Equation is at a loss to account for this finding since it makes no reference to a connection between antecedent and consequent. In contrast, according to inferentialism, the semantic core of HIT, such a connection should make a significant difference to how the probability of that conditional is evaluated.

In this paper, we test the implications of HIT for the probabilities of conditionals. While it is possible to assign probabilities to abstract conditionals (e.g., Evans et al., 2003), like the ones used in Douven et al. (2018), a more psychologically viable test draws on everyday conditionals—which is what we have done in the studies to be reported below. This has the additional advantage of extending HIT to everyday conditionals, since our previous work (Douven et al., 2018, 2020) only involved abstract conditionals.

But first: what *are* the implications of HIT for the probabilities of conditionals? We start at the *computational* level of analysis (the *what* of computation, in the manner of Marr, 1982). Given that probabilities are probabilities of *truth*, we only need to unpack the truth conditions that inferentialism assigns to a conditional to know what follows from that position, and thus from HIT, for the probability of a conditional. The probability of "If φ, ψ" is the probability that "If φ, ψ" is *true*, which is the probability that the truth conditions of "If φ, ψ" are realized, which is—according to inferentialism—the probability that there is a compelling argument from φ plus background knowledge to ψ, in the sense explained above.

As for the *algorithmic* (or *processing*) level of HIT, we first note that, in general, it is not a priori (in the colloquial sense of this expression) whether we can make a compelling case for ψ on the supposition of φ together with background knowledge. (We say "in general": if, for instance, we already hold either φ or ψ to be false, we are sure that we cannot make such a case.) For instance, the first author is confident that he can make a compelling case for the claim that FC Barcelona (a famous football club that is not doing so well at the moment) will be able to recover on the supposition that it buys some new top players, but he is not entirely certain of this. (The other authors have no opinion either way.) He would have to think more carefully about what players are available on the market, what internal difficulties the team may be facing, what rival teams are doing to increase their chances of success, and so on. His probability for "If Barcelona buys some new top players, they'll be able to recover" is his estimate that he can make the case, and in making that estimate he uses the heuristic of gauging the inferential strength between antecedent and consequent, that is to say, of gauging how strongly the consequent follows from the antecedent. In other words, we are proposing that inference strength translates into confidence, articulated as probability (Elqayam & Evans, 2013). We dub this heuristic process, temporarily, "the *inference heuristic*," and will discuss it further in the general discussion (Sect. 4).

The Ramsey Test and the New Paradigm. The New Paradigm is a family of theories rather than a single theory. We think of HIT as belonging to this family. We also share with New Paradigmers the decision-theoretic approach to reasoning (albeit with soft rather than strict Bayesianism; Elqayam & Evans, 2013), and more importantly, we share the theoretical commitment to the Ramsey Test, which refers to a famous footnote in Ramsey (1929/1990):

> When two people are arguing "If p will q?" and are both in doubt as to p, they are adding p hypothetically to their stock of knowledge and arguing on that basis about q; so that in a sense "If p, q" and "If p, ¬q" are contradictories. We can say they are fixing their degrees of belief in q given p. If p turns out false, these degrees of belief are rendered void. (p. 247)

In other words, to determine the degree to which you ought to believe a conditional, suppose its antecedent and, under that supposition, estimate how probable the conditional's consequent is.

The outcome of this procedure (or "test") yields at once your degree of belief for the conditional and your conditional degree of belief for the consequent given the antecedent. That, at any rate, is what the above footnote is standardly taken to imply.

Early on, the Ramsey Test became a foundation of the New Paradigm, with seminal papers (Evans et al., 2003; Over, Hadjichristidis, Evans, Handley, & Sloman, 2007) supporting interpreting it through the lens of the probability conditional. Evans and colleagues were the first to point out the startlingly psychological nature of the test, based as it is on processes such as hypothetical thinking (Evans, 2007a; Evans & Over, 2004). We concur; the Ramsey Test is every bit as important to HIT as it is to the probability conditional, and we agree that hypothetical thinking and heuristics are involved at the processing level. Where we differ is in the interpretation of the Ramsey Test itself. A look at the original quote shows that there is nothing whatsoever to link the Ramsey Test directly to the Equation, and indeed there is a long list of prominent philosophical contributions with alternative semantic interpretations of the Ramsey Test (Andreas & Günther, 2019, in press; Rott, in press; Stalnaker, 1984).

It was said that, in the FC Barcelona example, the first author estimates the probability of the relevant conditional by gauging the strength of the inferential connection between antecedent and consequent. That is the mental procedure he goes through in this case, a case we take to be entirely non-special, and which illustrates that the Ramsey Test is entirely compatible with inferentialism, and even more so with HIT: the hypothetical thinking involved requires the supposition of the antecedent plus background knowledge, from which the reasoner heuristically gauges whether it would be possible to make a compelling case for the consequent (*inference* heuristic). In previous work, we have started to map out this heuristic process. For example, in Douven et al. (2018) we showed that it uncannily mimics the classic pattern for belief bias, in which both deductive and heuristic parameters affect the response pattern.

The accumulation of evidence in favor of the Equation interpretation of the Ramsey Test is substantial; however, we argue that it is not entirely conclusive. In many cases, conditional probability would yield similar results to inference strength, creating a confound. Most of the literature supporting the Equation never controlled for this confound; as mentioned, when Skovgaard-Olsen (e.g., Skovgaard-Olsen et al., 2016; Skovgaard-Olsen, Singmann, & Klauer, 2017) did introduce controls, the effect of the Equation was only retained for conditionals whose antecedent was positively probabilistically relevant to their consequent. In the work we report here, we introduced more nuanced controls to examine in depth the relative roles of inference strength and conditional probability on estimates of the probability of conditionals.

Before we move on to the experiments, we recall that the principle of relevant inference postulates that the default mental representation of conditionals is one in which there is an inferential connection. This allows room for a non-default representation that is not governed by inference, although, being non-default, it would require more cognitive effort. Indeed, previous work (Douven et al., 2018, Exp. 2) found that, although inference strength is a reliable predictor of truth values of conditionals, it is not the sole predictor. We therefore predicted that inference strength, as measured in a separate task, should strongly predict the probability rating of a conditional, but also allowed some residual variance to be predicted by conditional probability, as a non-default mental representation. Still, we expected the predictive power of conditional probability, when controlling for inference strength, to be much lower.

In addition to this main hypothesis, we also manipulated three qualitatively different types of inferential links between conditionals: positive, negative, and missing link.[5] While inference

[5]Skovgaard-Olsen et al. (2017) measured a related 5-point scale of reason relations ranging from "a strong reason against," "a reason against," "neutral," and "a reason for" to "a strong reason for." Although this scale is a step in the

strength can be represented on a single scale, different sections of this spectrum are also qualitatively different in that they may be expected to lead to different categorical truth values. Hence, it was important to manipulate the three qualitatively different links, as well as the relative strength of the link. It also helped to disconfound the materials, since using only positive link materials risks a confound with conditional probability. As we have not used this type of manipulation before, we had an exploratory hypothesis concerning the difference between the three link types.

2 Experiment 1

Experiment 1 asked participants to judge, for a set of fifty conditionals, their probabilities, the strength of the inferential connection between their component parts, and the corresponding conditional probabilities. We were thus able to investigate the connection between probability and inference strength as implied by HIT as well as the connection between probability and conditional probability as stipulated by the suppositional account. Our goal was twofold: testing the hypothesis that how strongly people agree that ψ follows from φ predicts the probability they assign to "If φ, ψ," and answering the question of how the strength of people's agreement that ψ follows from φ compares, in terms of predictive accuracy, with their conditional probability for ψ given φ.

To achieve this twofold goal, we ran a study based on the method in Over et al. (2007), with two important changes: first, we added a task measuring inference strength; and second, to prevent participant fatigue and minimize potential for carry-over, we ran the study in three phases, about a week apart from each other. Each phase presented participants with a separate task.

2.1 Method

2.1.1 Participants

Participants were recruited by the INSEAD–Sorbonne University Behavioral Lab. From a total of 203 participants who completed the first phase, 154 participants completed all phases of the study. This high attrition rate (24 percent) is not unusual in longitudinal design, and importantly, the participants who completed the study did not differ in gender, age, or education level from the participants who dropped out. Of the 154 participants who completed all three phases, 36 were excluded for having had advanced training in logic, or a dyslexia diagnosis, or for not indicating French as their native language, or for failing an attention check.[6] This left us with 118 participants (82 women, 34 men, 2 non-binary/other gender individuals; 3 participants with a high school education and 113 with a post-secondary education; M_{age} = 22.51, SD_{age} = 3.48). Participants were

---

right direction, it does not allow to distinguish between inference strength and type of inferential link ("a strong reason against" and "a reason against" could plausibly measure negative link, "a strong reason for" and "a reason for" could measure positive link, while the single point "neutral" would most plausibly measure missing link). Our measure of inference strength is more direct and measures purely inference strength, leaving inference type as a (manipulated) independent variable.

[6]There were four attention checks in total: (1) In the demographics section, participants were given a list of film titles, and instructed to write "I read the instructions" in the "Other" box; this procedure was adapted from Pennycook, Trippas, Handley, and Thompson (2014). (2) In phase 1, participants were presented with a photograph of flamingos, and asked to count how many there were. (3) In phase 2, participants were presented with a drawing of three colored balls, and asked to indicate the color of the leftmost one. (4) At the end of the study (phase 3), participants were asked if they had answered seriously; this procedure was adapted from Aust, Diedenhofen, Ullrich, and Musch (2013). Participants were excluded if they failed any of the checks.

compensated through a lottery system, with eight randomly selected participants receiving € 40 each.

2.1.2 Materials and procedure

The study was run online using the Qualtrics platform, and consisted of three tasks presented in the following order: the probability task, the inference strength task, and the probabilistic truth-table task (all outlined below). The first and the last task were adopted from Over et al. (2007), whereas the second task was adopted (with necessary changes) from Douven et al. (2018). To avoid fatigue and carry-over effects, these tasks were separated from each other by a period of time ranging from five to seven days. All three tasks drew on a set of fifty conditionals which referred to possible real world events, such as, "If the French football team wins another World Cup, then football will become the most popular sport in France," and, "If a cure for AIDS is discovered, then condom sales will drop." All materials were presented in French. Some of the materials were adapted from Over et al. (2007), and some were created specifically for this study. The same materials were used for all three tasks.

Probability of the conditional task. Participants first completed the probability task of the study. In this task, participants were instructed to rate the probability of the fifty conditionals. Participants were instructed that some of the statements they would read would refer to a specific country and/or period, but that if no country or period was mentioned, they should assume that the statement referred to their country of origin within the next ten years. Probabilities were presented on a 0 to 100 percent slider scale, in which 0 percent meant that the statement was certainly false, and 100 percent meant that the statement was certainly true. Following a practice item, the fifty conditionals were presented in an individually randomized order.

Inference strength task. The URL for the inference strength task was sent to participants between five and seven days after completion of the probability task. In this second phase, participants provided ratings of inference strength. They were instructed to suppose a statement that corresponded to the antecedent of one of the conditionals used in the first part of the study, and they were then asked how strongly they agreed that another statement, which corresponded to the consequent from that same conditional, followed; for instance, participants were asked: "Suppose that a cure for AIDS is discovered. How strongly do you agree that it then follows that condom sales will drop?" Ratings were to be given on a 7-point Likert scale that ranged from "Strongly disagree" to "Strongly agree," with "Neither agree nor disagree" as the midpoint. Following a practice item, the fifty items were presented in an individually randomized order.

Probabilistic truth-table task. The URL with the probabilistic truth-table task was sent to participants between five and seven days after they completed the inference strength task. This third task required participants to rate, for each of the conditionals in our materials, the probability of four situations, which had the schematic form of $\varphi \,\&\, \psi$, $\varphi \,\&\, \neg\psi$, $\neg\varphi \,\&\, \psi$, and $\neg\varphi \,\&\, \neg\psi$, with φ the antecedent of the conditional and ψ the consequent. For instance, participants were shown the following table:

It is TRUE that a cure for AIDS will be discovered and it is also TRUE that condom sales will drop	. . .
It is TRUE that a cure for AIDS will be discovered but it is FALSE that condom sales will drop	. . .
It is FALSE that a cure for AIDS will be discovered but it is TRUE that condom sales will drop	. . .
It is FALSE that a cure for AIDS will be discovered and it is also FALSE that condom sales will drop	. . .
	100

Participants were asked to rate each situation on a probability scale ranging from 0 to 100 percent, in which 0 percent meant that the statement was certainly false, and 100 percent meant that the statement was certainly true. They were instructed that the probabilities should sum to 100 and were prevented from proceeding to the next item if this failed to be the case.

2.1.3 Data availability

For both experiments, the materials, data, and R files for the analysis are available at `https://osf.io/7x63j/?view_only=7db2e26d2d5b497faac451c531c800ff`.

2.2 Results

2.2.1 Analytic approach

To test our predictions regarding the effect of inference strength ratings and conditional probability ratings on probability judgements for a conditional, we fitted Bayesian regression models with the R package brms (Bürkner, 2018) and the probabilistic programming language Stan (Carpenter et al., 2017), which uses Markov Chain Monte Carlo algorithms. A Bayesian analysis estimates model parameters as probability distributions, with the joint probability distribution of the data, y, and a given parameter, θ, being computed via the prior probability of θ and the likelihood $p(y \mid \theta)$, as follows:

$$p(y, \theta) \quad = \quad p(y \mid \theta) \times p(\theta).$$

This result is derived from Bayes' Rule, which serves to calculate the posterior probability, $p(\theta \mid y)$, such that:

$$p(\theta \mid y) \quad \propto \quad p(y \mid \theta) \times p(\theta) \quad = \quad p(y, \theta).$$

This posterior probability distribution can be interpreted as indicating the relative plausibility of possible values of the parameter θ, conditioned on the prior probability of that parameter, the probability distribution of the responses (or likelihood function), and the data itself.

Given that responses to the probability of the conditional task were collected on a 0 to 100 percent continuous slider scale, we chose to model them as arising from a Gaussian distribution. We also specified prior distributions over the possible effects each parameter could have on the response variable. Specifying these priors is recommended because it allows regularization of parameter estimates (e.g., Bürkner, 2018; McElreath, 2020). For all models reported in this experiment, we specified weakly informative priors that constrain mean response estimates within the range of the response scale and indicate that extreme effects for the main predictors are unlikely while remaining agnostic to the direction of these effects.[7] Finally, because we used a repeated measures design where participants provided ratings for multiple items and where items were rated by multiple participants, we also include a (hierarchical) mixed-effects structure to our models, which estimates how group-level (or random) effects deviate from population-level (or main) effects and accounts for possible correlations in responses provided by the same participant or to the same item.

2.2.2 Model 1: Inference strength and conditional probability

We began by analyzing how judgments of the probability of a conditional depend on inference strength and conditional probability ratings. On the basis of the responses to the probabilistic

---

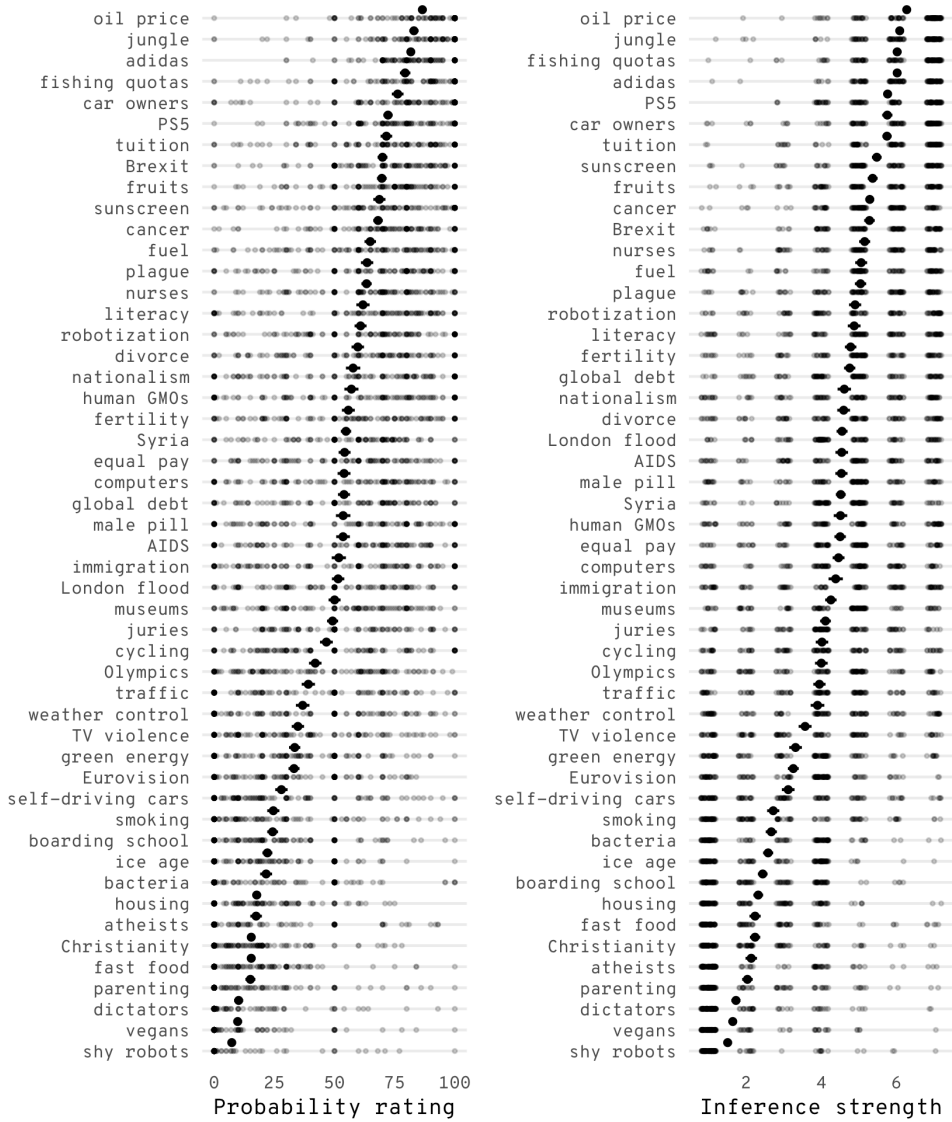[7] See Figure 9 in the Appendix for further details about priors specifications.

Figure 1: Probability ratings (left) and inference strength ratings (right) with horizontal jitter for ease of readability. Point and interval summaries represent the per item mean and standard error.

truth-table task, we derived conditional probability ratings, $p(\psi \mid \varphi)$, as in Over et al. (2007), by using the following formula:[8]

$$p(\psi \mid \varphi) \quad = \quad \frac{p(\varphi \,\&\, \psi)}{p(\varphi \,\&\, \psi) + p(\varphi \,\&\, \neg \psi)} \tag{1}$$

The mean conditional probability rating across all participants and conditionals was 65.3 (SD = 34.4) and the mean probability rating was 47.9 (SD = 33.7). As shown in Figure 1, our materials exhibited clear variation in probability ratings between different conditionals. The mean

---

[8]In the cases where it was not possible to directly derive conditional probability ratings with this formula (because $p(\varphi)$ was equal to zero), we estimated the conditional probability rating as being equal to 100.

inference strength rating across all participants and conditionals (the degree to which participants agreed that the conclusion followed from the premise) was 4.13 (SD = 2.07). Importantly, mean inference strength ratings per conditionals spanned the entire range of the scale (see Fig. 1), indicating that our materials included a wide variety of inference strength levels, which enabled us to investigate how conditionals that vary in the strength of their inferential connection might elicit different probability ratings.

Next, we fitted a Bayesian mixed-effects linear regression model, which regressed probability ratings on Inference strength and Conditional probability. Inference strength was specified as a monotonic predictor, given that responses to the inference strength task were collected on an ordinal Likert scale. Monotonic effects are the recommended approach for modelling predictors where one can expect a monotonic (increasing or decreasing) relationship between the predictor and the response but where intervals on the scale cannot be assumed to be equidistant and should therefore be allowed to have varying effects on the response (Bürkner & Charpentier, 2020). This model also included a mixed-effects hierarchical structure with two grouping factors (participants and items) over which random intercepts and random slopes for the two predictors of interest were also estimated:

Model 1: `Probability ~ mo(Inference strength) + Conditional probability +`
`(mo(Inference strength) + Conditional probability | Participant) +`
`(mo(Inference strength) + Conditional probability | Item)`

MCMC diagnostics indicated sufficient mixing of the chains, sufficiently high bulk and tail effective sample size values, and an $\widehat{R}$ convergence diagnostic of 1.00 for all parameters, which is within the recommended value range (Vehtari, Gelman, Simpson, Carpenter, & Bürkner, 2020).

Model 1 estimated a positive effect both of Inference strength (b = 7.97, 95 % CI [7.33 : 8.61]) and of Conditional probability (b = 0.10, 95 % CI [0.07 : 0.13]), indicating that probability ratings for a conditional increased by approximately 8 points for each point increase in Inference strength and by approximately 0.1 points for each point increase in Conditional probability. Given that ratings for these two variables were collected on different scales, the difference in size
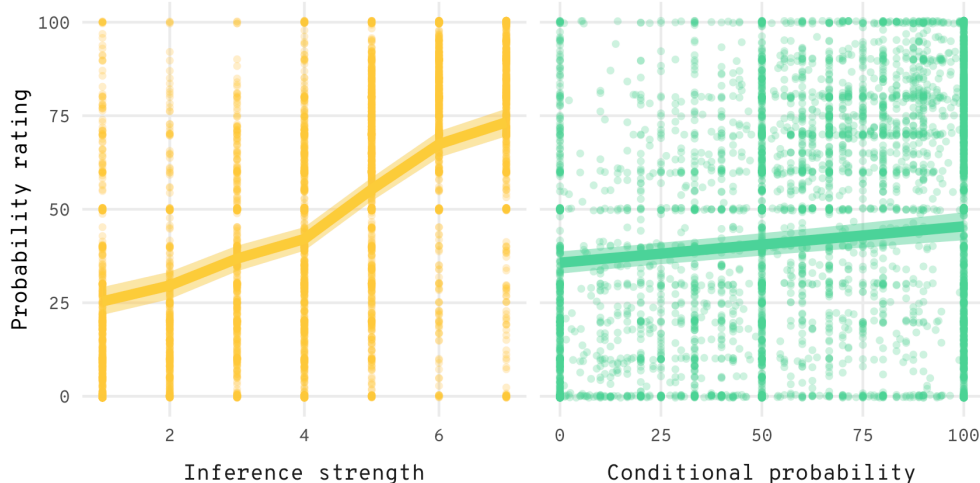

Figure 2: Conditional effects plots for the Inference strength (left) and Conditional probability (right) predictors, with 95 percent compatibility intervals and data overlayed. Effects for each predictor were estimated conditional on the mean value of the other predictor.
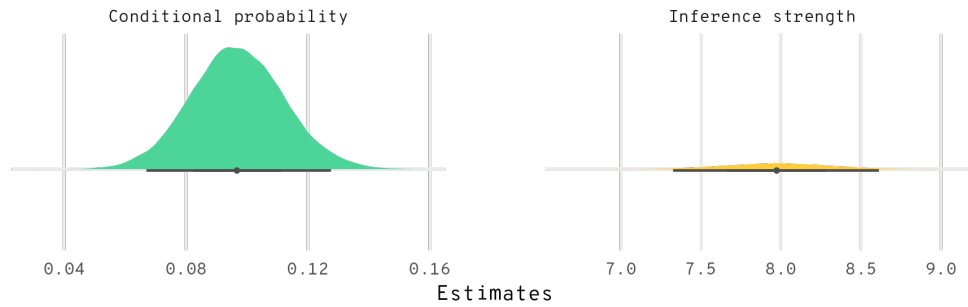
Figure 3: Posterior probability distributions (with mean and 95% CI) of the parameter estimates for Inference strength and for Conditional probability.

between the estimates is difficult to interpret directly. However, as is apparent in the conditional effects plot shown in Figure 2, the estimated slope for the Inference strength predictor was clearly steeper than the estimated slope for the Conditional probability predictor. Notably, when conditioning on Inference strength at its mean level, an increase in Conditional probability from 0 to 100 (the full range of the scale) only predicted an increase in probability ratings of approximately 20 points (on a 0 to 100 scale), whereas, when conditioning on Conditional probability at its mean value, an increase in Inference strength from 1 to 7 (again, the full range of the scale) predicted an increase in probability ratings of over 50 points.

We can summarize the results from this model by inspecting the posterior probability distribution of the model parameters. Figure 3 shows that plausible parameter values for Inference strength are distributed rather uniformly while being consistently superior to 7.5. This can be interpreted as indicating that Inference strength has a clearly positive, and relatively strong, effect on probability ratings, regardless of the second predictor that is included in the model. The distribution of plausible parameter values for the probability measure has a much narrower spread and is centered on comparatively much lower values (note the difference in the scale of the *x*-axis between the two panels), although it remains mostly positive. This suggests that in Model 1 probabilistic considerations have a meaningful, albeit small, impact on judgments of the probability of a conditional.

2.2.3 Individual differences analysis

We followed Over et al. (2007) once more by also running an analysis of individual differences. Earlier studies on the probabilities of conditionals (e.g., Evans, Handley, Neilens, & Over, 2007; Fugard et al., 2011; Over et al., 2007) found a consistent pattern of individual differences, in which the majority of participants responded according to a conditional probability pattern, with a large minority adopting a conjunctive pattern; however, the latter disappeared as participants accrued practice (Fugard et al., 2011) or when the materials were not abstract (Over et al., 2007). For example, Over et al. (2007) compared as predictors of people's probability assignments to conditionals, both conditional probabilities and conjunctive probabilities—that is, for "If φ, ψ," the probability of the conjunction of φ and ψ. Importantly, however, these earlier studies did not control for inference strength. In more recent work, Skovgaard-Olsen, Kellen, Hahn, and Klauer (2019) used a choice paradigm in which participants were instructed to decide between fictional protagonists advocating either an inferentialist approach or a conditional probability approach. They found that a majority of participants preferred the inferentialist approach, and

that their choices predicted their truth assignment for conditionals. A large minority conformed to a conditional probability approach, and again this predicted their truth assignments. However, this study drew on explicit choice, and so would not have been able to identify any implicit patterns. None of these studies tested individual differences predicting the probability of conditionals while accounting both for probability measures (such as conditional probability) and inference strength. This is what our analysis does.

Instead of directly replicating the method from Over et al. (2007), who performed individual (per participant) regression analyses across their items, with the probability of the conditionals as the dependent variable and their other variables of interest as predictors, we directly extracted from our two models, as well as from a second model (Model IS + CJ) that replaces conditional probabilities with probabilities of conjunctions, the group-level (or random) effects estimated for each participant. These group-level effects take the form of group-level deviates, which the statistical model assumes to be normally distributed around the population-level effect. Because Bayesian inference works with samples from the posterior distribution of the estimated parameters, we are able to directly compute the estimated effect of the predictors of interest for each participant by adding, for each sample in the posterior, the estimated population-level effect of each predictor with the corresponding estimated group-level deviate.

Figure 4 indicates that group-level estimates for the Inference strength predictor are consistently positive, with the exception of one participant. These group-level estimates are superior in most cases to a mean estimated coefficient of 5, meaning that, for most participants, the model predicted that a one point increase in Inference strength should lead to at least a five point increase in probability ratings, and in some cases more than a ten point increase in probability ratings. The figure also shows a relatively high amount of between participant variation, suggesting that some participants were more sensitive to inferential strength than others in their interpretation of conditional statements. Group-level estimates for the probabilistic measures also exhibit a certain amount of between participant variation, but they appear constrained to a very narrow range. They are consistently very close to 0 (viz., lower than 0.4) and in a sizeable minority of cases for both Conditional and Conjunctive probability they are compatible with null and even negative parameter values. This suggests that the estimated contribution of probability measures to probability ratings corresponded at best for participants to less than half a point increase in probability ratings for each point increase in one of the probability measures.[9]

2.2.4 Cross-validation results

Finally, we cross-validated our models using the loo package (Vehtari, Gabry, et al., 2020) which performs leave-one-out cross-validation using Pareto smoothed importance sampling, or PSIS-LOO (Vehtari, Gelman, & Gabry, 2017). This method provides estimates of the point-wise out-of-sample prediction accuracy (or ELPD) of a model, as well as approximations of the standard error for the estimated prediction error of a model, thereby enabling comparisons in predictive accuracy between models. We compare Model 1 with restricted models corresponding to simpler hypotheses (e.g., the hypothesis that Inference strength predicts probability ratings while Conditional probability has a null effect). This verifies that the model embodying our hypothesis is

---

[9]Following a recommendation from the editor, we also performed individual (per participant) regression analyses across the fifty items, where a fixed effects version of the three models was fitted for each participant separately. The effects estimated by this approach showed a larger amount of variation and of interval uncertainty because of the lack of shrinkage provided by a random effects model but were qualitatively very similar. See Figure 11 in the Appendix for the corresponding results.
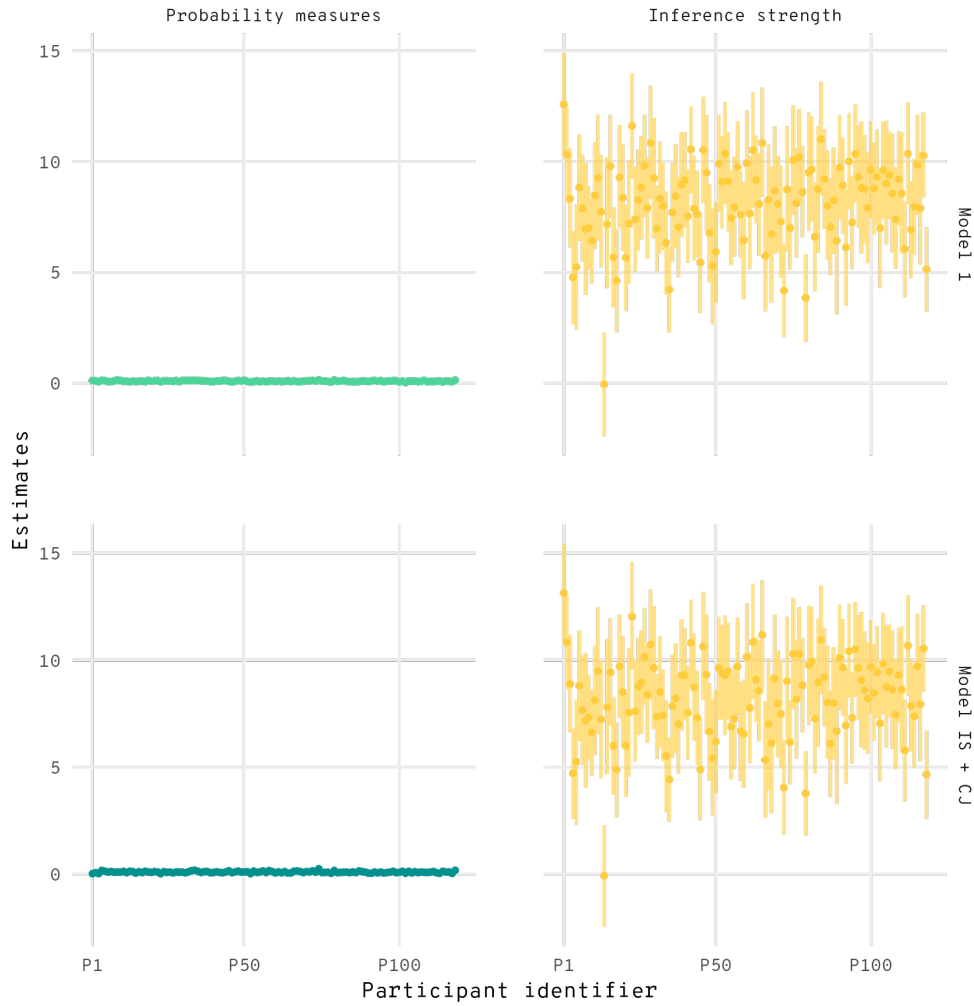
Figure 4: Mean posterior probability (with 95 % CI) for the group-level parameter estimates per participant for Inference strength and for two probability measures: Conditional probability (Model 1, in light green), and Conjunctive probability (Model IS + CJ, in dark green). (NB: Probabilities were measured on a scale from 0 to 100; inference strength was measured on a 7-point Likert scale.)

not over-fitting the data (i.e., fitting variation in the data that is simply due to random noise) and that our hypothesis does not introduce complexity that is unwarranted by the data.

Table 1 summarizes the comparisons in predictive accuracy between the different models. These comparisons were done by computing the difference in ELPD (Δ_{ELPD}) between each model and the model with the largest ELPD and then by verifying that the standard error of this difference (Δ_{ELPD} SE) is several times smaller than the difference itself. Models IS and CP are restricted versions of Model 1, which only include one predictor both as population-level and group-level effects (Inference strength and Conditional probability, respectively).

As reported in Table 1, the difference in ELPD between the two simpler models (Models IS and CP) and Model 1 (identified as the best model) is several times larger (four and twenty times, respectively) than the standard error of that difference, indicating that they both have worse predictive performance than Model 1. This suggests that including both predictors improves model predictions meaningfully.

Table 1: Cross-validation comparisons of the models in Experiment 1.

Model	Δ_{ELPD}	(SE)
1	**0.0**	**0.0**
IS	−52.4	13.5
CP	−1035.7	54.6

Note: In a cross-validation comparison, the Δ_{ELPD} is the difference in ELPD between each model and the model with the largest ELPD (indicated in bold face). The Δ_{ELPD} for the best model is always 0 given that it is the difference between that model and itself. SE corresponds to the standard error of the difference in ELPD between two models and indicates the amount of uncertainty present in the comparison. The Δ_{ELPD} must be several times larger than the SE for the comparison to indicate meaningful differences in predictive performance between two models.

2.3   Discussion

Our goal was to investigate whether judgments of the probability of a conditional are better predicted by inference strength ratings or by probabilistic measures, most notably, conditional probability ratings, which we derived from the probabilistic truth-table task. Our hypothesis was that the strength of the inferential connection would be the main predictor of the probability of a conditional because it corresponds to the default mental representation of conditional statements. It also left open the possibility that non-default representations could be activated when interpreting conditionals, allowing for probabilistic measures to also contribute, to some extent, as predictors of probability ratings.

Modeling probability ratings as predicted both by inference strength and conditional probability revealed that inference strength was a stronger predictor than the probabilistic measure (even when accounting for the fact that the scale of this probabilistic measure had a much wider range than inference strength). A similar pattern appeared when analyzing predictor estimates at the individual level. When the effect of conditional probabilities was examined, Model 1 estimates indicated that, after controlling for the effect of inference strength, conditional probability was a weak but meaningful predictor of probability ratings. At the individual level, mean estimates of the group-level effect of conditional probability were positive for all participants, with only a small range of the 95 percent CI for some participants indicating that negative or null values for that predictor were also plausible given the data. Finally, a cross-validation analysis of Model 1 indicated that both Inference strength and Conditional probability contributed to improving meaningfully the predictive performance of that model. These results provide evidence in favor of the two parts of our hypothesis. They suggest, first, that inference strength has a strong and meaningful influence on probability ratings, and second, that conditional probability has a small, but overall reliable, influence on probability ratings as well. This supports our proposal that strength of the inferential connection and conditional probabilities correspond to parallel representations of conditional statements, which are activated to varying degrees depending on individuals, with the inferential representation being possibly the default one given that it appears to drive probability ratings more strongly. Notably, however, our individual differences analysis found that there were almost no cases where conditional probabilities had a larger effect on probability ratings than inference strength, and no cases where conditional probability ratings had a large absolute effect.

This experiment had three important limitations. First, because inference strength responses and probability ratings had been collected on different scales, direct comparisons of the size of

the predictors were difficult, and one could raise the concern that the predictive success of inference strength might be an artefact of a scale that is simpler to interpret. Second, and relatedly, we derived conditional probability ratings from the probabilistic truth-table task, a task which, inter alia, required participants to ensure that their responses summed to 100. This task was therefore cognitively more demanding than the other tasks and it was also less direct and less intuitive than the judgment which it purports to measure, to wit, the probability of the consequent on the assumption that the antecedent is true. The first two limitations might explain, in part, why, contrary to previous results in the literature, we did not find meaningful evidence in favor of a mental representation where conditional probabilities were the dominant predictor of probability ratings. Finally, although Figure 1 underlines the large amount of between item variation in inference strength, this experiment did not explicitly include conditionals with different types of inferential link. In particular, it did not allow us to investigate how inference strength and conditional probabilities fare when predicting the probability ratings of missing link conditionals. We address these limitations in Experiment 2.

3 Experiment 2

The second experiment followed closely the design of Experiment 1. We sought to further test our proposal that the probability of a conditional is predicted both by the strength of the inferential connection between its component parts and, to a lesser extent, by probabilistic considerations (measured as conditional probabilities). We also sought to compare our proposal with the suppositional account, according to which conditional probabilities are the sole predictor of probability judgments for a conditional. This experiment improved upon Experiment 1 by using a simpler and more direct measure of conditional probability judgments and by including three different types of conditional statements, to be defined below as "positive link," "missing link," and "negative link" conditionals.

The inclusion of missing link conditionals presents an especially interesting test case for the comparison of inferentialism and the suppositional account, given that missing link conditionals are characterized by a weak or even absent inferential connection but can have any—and so also a high—conditional probability. This means that, given inferentialism, missing link conditionals are expected to receive generally low probability ratings whereas their probability ratings could be anything, given the suppositional account.

3.1 Method

3.1.1 Participants

Participants were again recruited by the INSEAD–Sorbonne University Behavioral Lab. From a total of 272 participants who took part in the first phase of the experiment, 244 participants completed all three phases, corresponding to a 10 percent attrition rate. This attrition rate is, again, not unusual in a longitudinal design and participants who completed the experiment did not differ meaningfully in gender, age, or education level from the participants who dropped out. Applying again the same criteria as those described in Experiment 1 (excluding participants who had received advanced training in logic, a dyslexia diagnosis, who had indicated that French was not their native language or had failed any of the four attention checks) led to the further exclusion of 40 participants, which left us with 204 participants (155 women, 45 men, 4 nonbinary/other gender individuals; 6 participants with a high school education and 198 with a post-

secondary education; M_{age} = 22.7, SD_{age} = 3.63) for the analyses. Participants were compensated through a lottery system, with eight randomly selected participants receiving € 40 each.

3.1.2 Materials and procedure

This second experiment followed a procedure almost identical to the one described for Experiment 1, with three notable differences. First, and in order to compare more directly the behavior of the inferentialist and probabilistic models in the case of missing link conditionals, this experiment used an updated set of fifty concrete causal conditionals which were classified as belonging to one of three categories: "positive link" conditionals, where there is a more or less strong inferential relationship between antecedent and consequent (e.g., "If a cure for AIDS is discovered, condom sales will decrease"); "missing link" conditionals, where there is no inferential relationship between the antecedent and consequent (e.g., "If bacteria become resistant to antibiotics, electric scooters will be forbidden in London"); and "negative link" conditionals, where one would rather infer the *negation* of the consequent from the antecedent than the consequent itself (e.g., "If treatments for infertility improve, the growth of the world population will slow down"). The set of conditionals presented to participants included 30 positive link, 10 missing link, and 10 negative link conditionals. This classification was developed on the basis of the authors' subjective interpretation of the meaning of these conditionals and was not shared with participants.

Second, the probabilistic truth-table task used in the third phase of Experiment 1 was replaced by a conditional probabilities task (described in more detail below). And third, the responses to the probability of the conditional task and to the inference strength task were collected on 11-point Likert scales (ranging from "Very unlikely" to "Very likely," with "Neither likely nor unlikely" as the midpoint, for the former, and from "Strongly disagree" to "Strongly agree," with "Neither agree nor disagree" as the midpoint, for the latter), and the prompt used for the inference strength task was simplified to "Does it follow that condom sales will drop?" (to reuse the example from Experiment 1).

Conditional probabilities task. This task took the place of the probabilistic truth-table task which was used in Experiment 1, and it sought to mirror more closely the psychological procedure which, according to the suppositional account, underlies the evaluation of the probability of a conditional. In this task, participants rated, for each of the conditionals in our materials, three probabilities. First, they were asked to rate the probability of the consequent of the relevant conditional, $p(\psi)$. After they had responded to that first question, a second question appeared on the same page, which asked them to suppose that the antecedent of the relevant conditional was true and then to rate the probability of the consequent under that supposition, $p(\psi \mid \varphi)$. Finally, after they had responded to the second question, a third question appeared, again on the same page, which asked them to suppose that the *negation* of the antecedent of the relevant conditional was true and then to rate, under that new supposition, the probability of the consequent, $p(\psi \mid \neg\varphi)$. The second question was meant to directly measure conditional probability judgments. For instance, participants were shown, one by one, the following series of questions:

$p(\psi)$:   In your opinion, how likely is it that condom sales will drop?

$p(\psi \mid \varphi)$:   Now suppose that a cure for AIDS will be discovered. In your opinion, how likely is it that condom sales will drop?

$p(\psi \mid \neg\varphi)$:   Finally, suppose that a cure for AIDS will *not* be discovered. In your opinion, how likely is it that condom sales will drop?

As mentioned, responses were collected on an 11-point Likert scale ranging from "Very unlikely" to "Very likely," with "Neither likely nor unlikely" as the midpoint.

3.2 Results

3.2.1 Analytic approach

Save for one key difference, we used the same analytic approach described for Experiment 1, specifying weakly informative priors (see Fig. 10 in the Appendix for more details) and including a mixed-effects hierarchical structure into our models. Because responses were collected on an 11-point Likert scale, we chose to analyze the data using an ordinal logistic regression model. An ordinal regression model assumes responses to have resulted from the categorization of a latent continuous variable (e.g., the strength of the inferential connection in a conditional) which is divided by respondents into bins (corresponding, for instance, to each point on a Likert scale) of possibly varying sizes (Bürkner & Charpentier, 2020). Importantly, the distance between two points on a Likert scale cannot be assumed to be the same for all pairs of contiguous points on the scale nor can it be assumed to be the same for all participants, two features which are incompatible with the assumptions of linear regression and which ordinal regression is specifically meant to account for (see e.g., Bürkner & Charpentier, 2020; Bürkner & Vuorre, 2019; Liddell & Kruschke, 2018).

The models reported in this experiment estimate how the explanatory variables influence the logit-transformed probability of respondents selecting a given response point on the response scale. The logit-transformation converts a probability p (which is, by definition, restricted to the 0 to 1 range) into a log odds ratio by taking the logarithm of the ratio between p and $1 - p$. A log odds ratio of 0 means that p and $1 - p$ are equal, a positive log odds ratio means that p is larger than $1 - p$, and a negative log odds ratio means that p is smaller than $1 - p$.

3.2.2 Model 2: Inference strength and conditional probability

We first examine how judgments of the probability of a conditional are influenced by inference strength and conditional probability ratings. The mean probability rating across all participants and conditionals was 5.51 (SD = 3.50), with the conditionals in our sample displaying a relatively large amount of between item variation (see Fig. 5). The mean conditional probability rating across all participants and conditionals was 6.93 (SD = 2.99) and the mean inference strength rating (the degree to which participants agreed that the conclusion followed from the premise) was 5.32 (SD = 3.53). Notably, as shown in Figure 5, participants' ratings broadly reflected our a priori classification of the conditionals included in the materials, with mean inference ratings per conditional being generally superior in the positive link condition than in the other two conditions.

The data displayed in Figure 6 exhibit a number of notable patterns. In the Positive link condition, probability ratings appear to be positively correlated with both inference strength and conditional probability ratings, with the upper quadrant of the two respective panels in Figure 6 containing a high density of data points. In the Negative and Missing link conditions, inference strength ratings and probability ratings are mostly distributed on the lower half of their respective scales. On the other hand, conditional probability ratings are distributed across the whole range of the scale, indicating that negative and missing link conditionals were likely to receive even high conditional probability ratings while rarely receiving probability ratings above the midpoint of the scale.
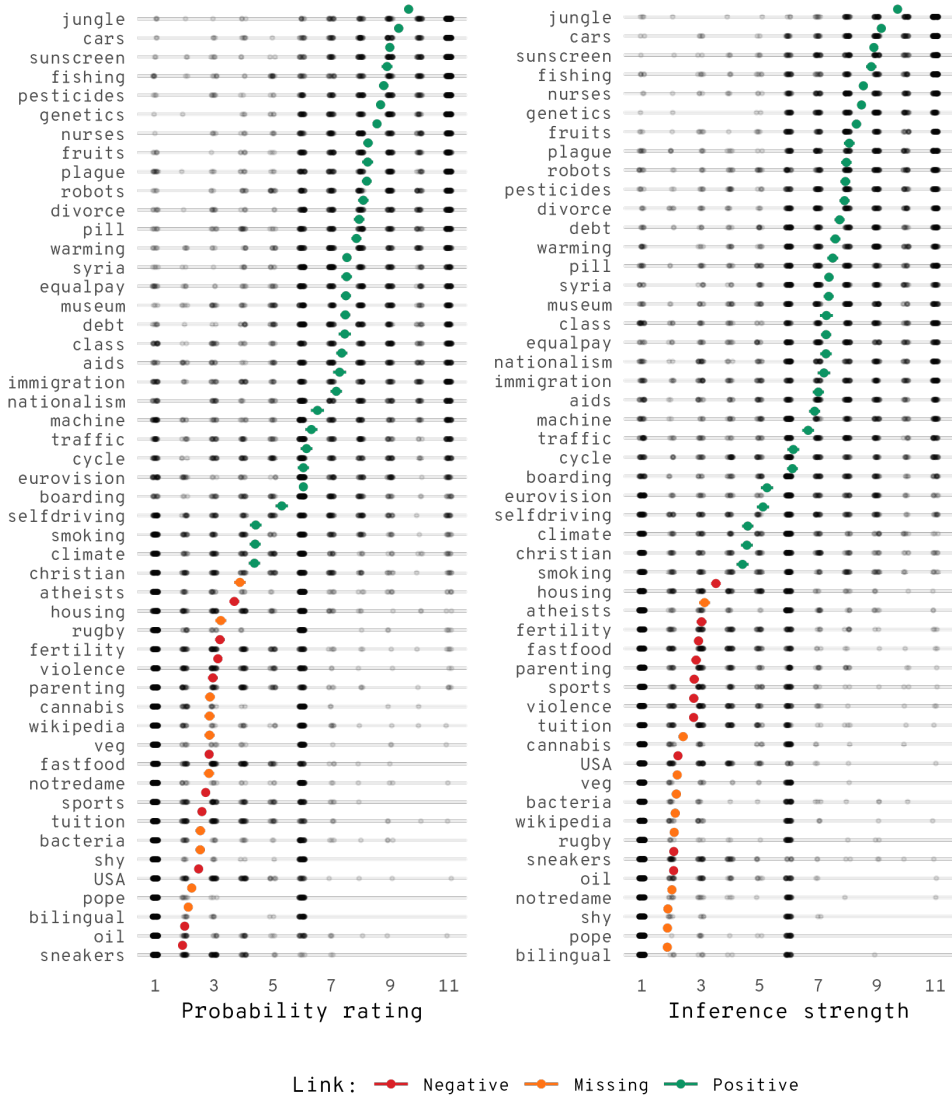
Figure 5: Probability ratings (left) and inference strength ratings (right). Data points are jittered horizontally for ease of readability. Point and interval summaries represent the per item mean and standard error and are color-coded according to their inferential-link type.

We analyzed the data by fitting a Bayesian mixed-effects ordinal logistic regression model, which predicted probability ratings using inference strength and conditional probability (both specified as monotonic predictors) as explanatory variables. These two predictors were each set to interact with Link (reference level: Positive), which allowed their effect on probability ratings to vary depending on the (a priori defined) type of inferential link. Note that this interaction effect is not a predicted consequence of HIT: negative and missing link conditionals simply present us with cases in which we can expect inference strength to be low, and therefore we expect these conditionals' probabilities to be low as well. However, we chose to include this interaction term in our models to allow us to assess whether the effect of inference strength and conditional probability varies depending on the type of conditional link. (See Table 2 for cross-validation results

confirming that this additional predictor did not lead to over-fitting in our models.) This model also had a mixed-effects hierarchical structure with two grouping factors (participants and items) over which group-level effects were estimated for the two predictors of interest and their respective interaction with type of inferential link. This model was defined as follows:

Model 2: `Probability ~ mo(Inference strength)*Link +`
`    mo(Conditional probability)*Link + (mo(Inference strength)*Link +`
`    mo(Conditional probability)*Link | Participant) +`
`    (mo(Inference strength)*Link + mo(Conditional probability)*Link | Item)`

MCMC diagnostics reported that the chains had sufficiently mixed, that bulk and tail effective sample size values were above recommended values, as well as an $\widehat{R}$ convergence diagnostic of 1.00 for all parameters.

The results from Model 2 confirmed our impression based on visual inspection of the data by indicating that higher probability ratings were less likely in the Negative link (b = –1.10, 95 % CI [–1.68 : –0.30]) and, with somewhat more uncertainty, in the Missing link condition (b = –1.02, 95 % CI [–1.94 : 0.37]) than in the Positive link condition. This model also estimated a positive effect of Inference strength (b = 0.50, 95 % CI [0.46 : 0.53]) and a positive, but relatively lower, effect of Conditional probability (b = 0.19, 95 % CI [0.14 : 0.24]) in the Positive link condition, meaning that increases in those predictors increased the probability of higher probability ratings. According to Model 2, the effect of Conditional probability was comparatively lower in the Negative link (b = –0.13, 95 % CI [–0.26 : –0.01]) and Missing link (b = –0.16, 95 % CI [–0.26 : –0.06]) conditions than in the Positive link condition. The estimated effect of Inference strength was also comparatively lower in the Negative link (b = –0.28, 95 % CI [–0.51 : –0.04]) condition. In the Missing link condition, this effect was estimated to be comparatively higher (b = 0.08, 95 % CI [–0.05 : 0.22]), but the estimated mean effect was close to 0 and the 95 percent compatibility interval included a small range of negative values, therefore indicating some uncertainty about the direction or even presence of the effect.

Given that estimates in log odds ratio are difficult to interpret intuitively, we also refer the reader to Figure 6, which shows the predicted relationship between our predictors of interest and probability ratings. In particular, this figure indicates that the estimated slope for Inference strength is positive in all three conditions and is somewhat less steep in the Negative link condition. On the other hand, the estimated slope for Conditional probability is almost flat in the Negative and Missing link conditions and is weakly increasing in the Positive link condition. Notably, in the Positive link condition, an increase in Conditional probability from 1 to 11 (the full range of the scale) only predicts an increase in mean probability rating of 2 points, while an equivalent increase in Inference strength predicts an increase in mean probability ratings of 4.5 points.

The results from Model 2 are summarized in Figure 7, which reveals that plausible parameter values for the Inference strength predictor are consistently higher than Conditional probability in the Positive and Missing link conditions, indicating a stronger effect of Inference strength on probability ratings in these two conditions.

3.2.3 Individual differences analysis

We analyze the effect of inference strength and probability measures at the individual level, by extracting the per participant group-level effects from Model 2, using the same procedure as described in Experiment 1. Because in the present experiment, the explanatory variables are on similar scales, we are able to directly compare them by computing for each participant the difference in
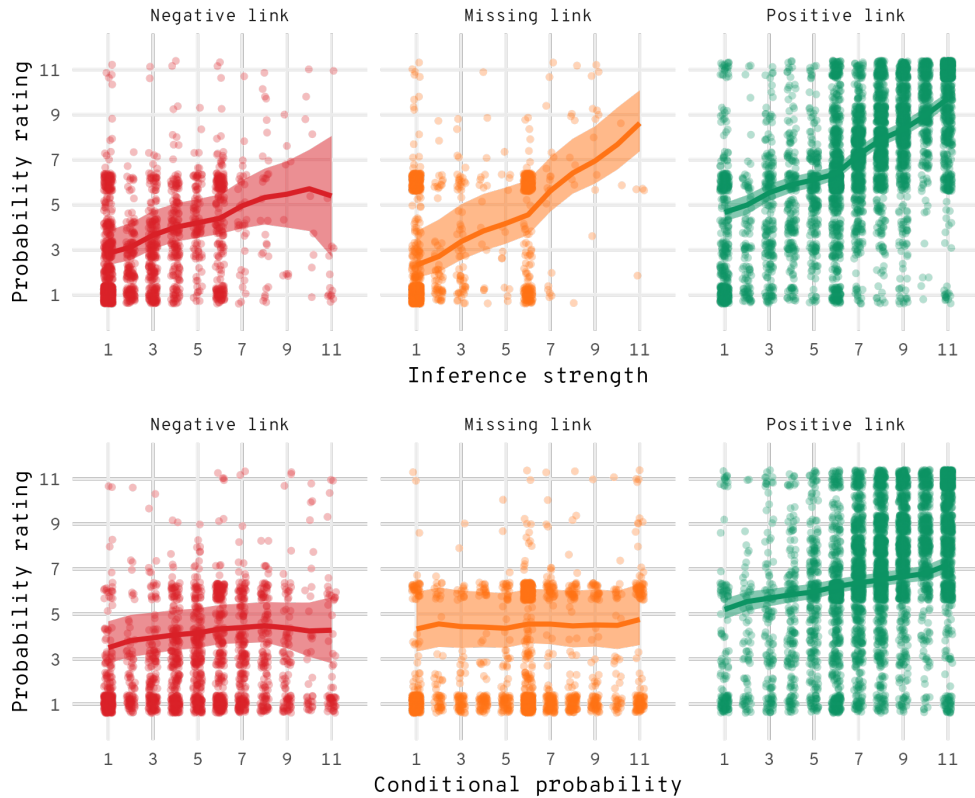
Figure 6: Conditional effects plots per type of inferential link for the Inference strength (top) and Conditional probability (bottom) predictors, with 95 percent compatibility intervals and data overlayed. Effects for each predictor were estimated conditional on the mean value of the other predictor.

estimated effects between the Inference strength estimate and the estimates for Conditional probability: a positive difference indicates that Inference strength had a stronger effect on probability ratings than the Conditional probability. Figure 8 shows that in the Positive and Missing link conditions, the mean differences in estimates are consistently superior to 0, while in the Negative link condition, mean differences in estimates exhibit more variation and can be either positive or negative.

To facilitate the interpretation of this figure, we propose to identify Inference strength as the dominant predictor when the mean difference in estimated effects per participant is greater than 0.18 in log odds ratio (this difference corresponds approximately to a Cohen's d of 0.2, which is considered to be a small effect size) and the probability measures to be the dominant predictors when this difference is smaller than −0.18 in log odds ratio. In the remaining cases, we do not identify either predictor as being dominant. We find that, following this proposal, a majority of participants are classified in the Positive link condition as following an inferential strength pattern (87 percent), with no participant being classified as following a probabilistic pattern. In the Missing link condition as well, a large proportion of participants (viz., 95 percent) are classified as following an inferential pattern, and the remainder are classified as not exhibiting a dominant reasoning pattern. Finally, in the Negative link condition, 41 percent of participants are identified as reasoning under an inferential strength and a very small proportion of participants (1 percent) are identified as reasoning under a probabilistic pattern.
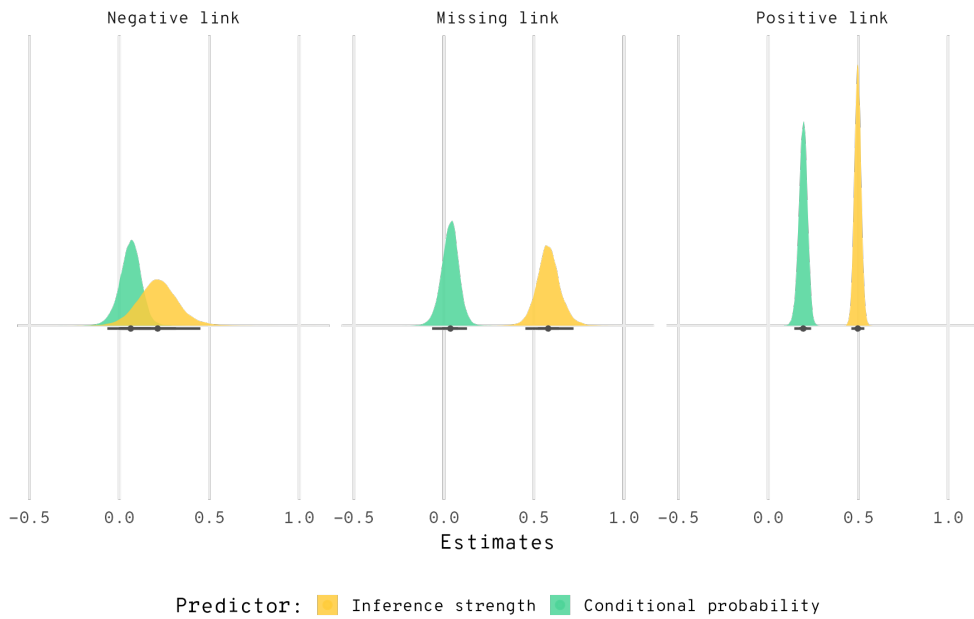
Figure 7: Posterior probability distributions (with mean and 95 % CI) of the parameter estimates per type of inferential link for Inference strength and for Conditional probability (Model 2).
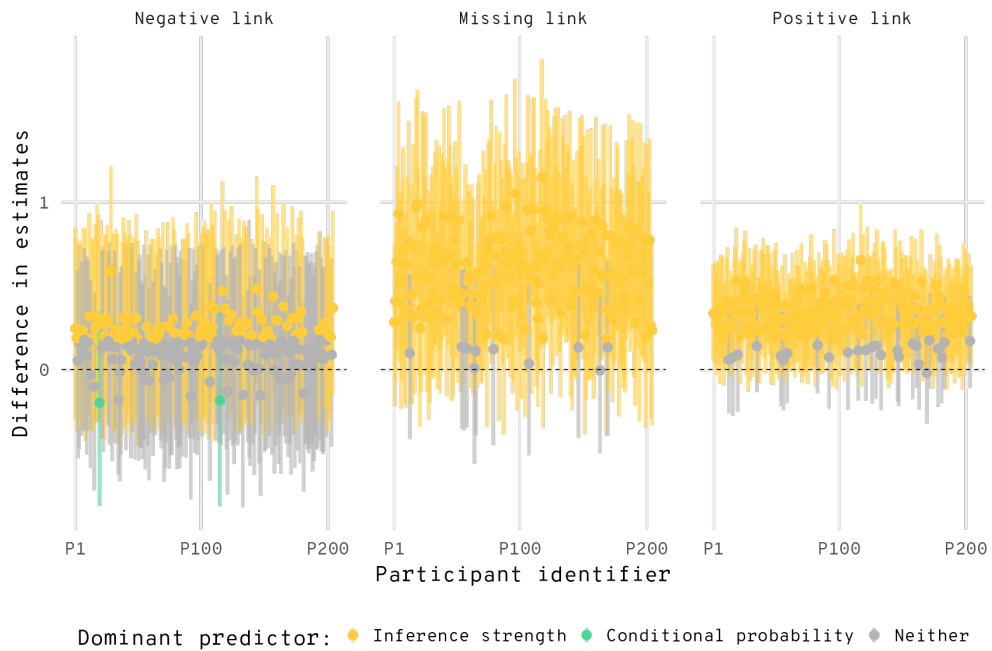

Figure 8: Mean difference in posterior probability (with 95 % CI) between the group-level parameter estimates of Inference strength and Conditional probability. A predictor is described as "dominant" when the mean difference between the two predictors is greater than 0.18, or, respectively, smaller than −0.18 in log odds ratio (a difference which corresponds, approximately, to a Cohen's *d* of 0.2).

We tested the sensitivity of this procedure by comparing it with both a stricter classification procedure (according to which a predictor was identified as dominant if the lower/upper boundary of the 95 percent CI was above/below the 0.18/−0.18 threshold), and a looser classification procedure which simply compared the mean difference in estimated effects per participant with a threshold value of 0. Under the looser procedure, Inference strength was the dominant predictor for over 99 percent of participants in the Positive link and Missing link condition and for approximately 88 percent of participants in the Negative link condition, with Conditional probability being the dominant predictor for the remaining participants. Under the stricter procedure, Conditional probability was not identified as dominant in any of the conditions and Inference strength was identified as dominant in 9 percent of cases in the Positive link condition, in 29 percent of cases in the Missing link condition and in none of the cases in the Negative link condition. In short, regardless of the classification procedure, Conditional probability was never identified as a dominant predictor for more than a minority of participants in any of the conditions, while Inference strength corresponded either to the dominant predictor or to a predictor equivalent to Conditional probability. [10]

3.2.4 Cross-validation results

We conclude our analysis by cross-validating our model of interest. The column labeled "CV 1" in Table 2 shows the results of a comparison of Model 2 with restricted models that only include, both as population-level and group-level effects, one of the predictors of interest and its interaction with the type of inferential Link: Models IS × L, and CP × L, corresponding to models regressing probability ratings on Inference strength and Conditional probability respectively, and their interaction with Link. The difference in ELPD between Models IS × L and CP × L on the one hand, and Model 2 (the best model) on the other hand, is multiple times larger (namely, nine and sixteen times larger) than the standard error of that difference, which indicates that Model 2 has better predictive performance than models that only include Inference strength or Conditional probability.

Next, we confirm that including Link as an additional predictor and interaction term in our model did not lead to over-fitting by comparing Model 2 with a model (Model ISCP) that does not include Link as a predictor; see the column labeled "CV 2" in Table 2. Model 2 is seen to have better predictive performance than the smaller model, with the difference in ELPD being twelve times larger than the standard error.

3.3 Discussion

In this experiment, we sought to investigate whether the findings of Experiment 1 regarding the relative influence of inference strength and conditional probability on judgments of the probability of a conditional would be replicated in a setting that used a more direct and intuitive measure of conditional probability than the one provided by the probabilistic truth-table task and that collected ratings of inference strength and conditional probability on scales with the same range.

We found very similar patterns of results as in Experiment 1. Specifically, inference strength had a positive impact on probability ratings of conditionals, and a stronger impact than condi-

---

[10] Here, too, we also performed individual (per participant) regression analyses across the fifty items, where a fixed effects version of both models was fitted for each participant separately. The effects estimated by this approach showed again a comparatively larger amount of variation because of the absence of regularization toward the population mean that is obtained with a random effects model. The alternative analysis did not reveal any notable differences from the random effects approach. See Figure 12 in the Appendix.

Table 2: Cross-validation comparisons of the models in Experiment 2.

Model	CV 1		CV 2	
	Δ_{ELPD}	(SE)	Δ_{ELPD}	(SE)
2	0.0	0.0	0.0	0.0
IS × L	−146.6	21.1	—	—
CP × L	−919.6	54.8	—	—
ISCP	—	—	−298.8	25.5

Note: For explanation, see the note to Table 1.

tional probability, which also had a positive impact on probability ratings. At the individual level, inference strength was the dominant predictor for a large majority of participants.

These results can be interpreted as providing further support for our proposal that the default representation of conditionals is an inferential one, but it also tentatively suggests the existence of a mixed regime, in which both the inferential connection and conditional probability play a role in the interpretation of a conditional.

A second aim of this experiment was to investigate how well conditional probabilities would be able to predict probability ratings for a conditional in cases where an inferential connection between antecedent and consequent was absent. We mentioned above that the Missing link condition constitutes a case where Inferentialism and the suppositional account make markedly different predictions, given that missing link conditionals lack an inferential connection but can have any—and so also a high—corresponding conditional probability. Interestingly, the conditionals in the Missing link condition received ratings that were in line with this characterization: their inference strength ratings were systematically low (with a majority of the data being clustered on the lowest and middle points of the scale) while their conditional probability ratings were evenly distributed across the entire range of the scale. If inferential considerations do not matter in the interpretation of missing link conditionals, one would therefore expect their probability ratings to be simply correlated with their conditional probabilities. If, however, inferential considerations do matter, then one would expect perceived inferential strength and not conditional probability to be better correlated with judgments of the probability of a missing link conditional. The results from Model 2 were clearly in favor of the inferentialist proposal: inference strength was strongly correlated with probability ratings while conditional probabilities were very weakly correlated with probability ratings. A similar pattern of results in the individual differences analysis lent additional support for inferentialism.

Our sample of materials also gave rise to a Negative link condition, covering conditionals from whose antecedent one would rather infer the negation of the consequent than the consequent. The results for this condition were mixed: although conditional effects plots showed generally similar patterns to the Missing link condition, inference strength estimates did not differ meaningfully from conditional probability estimates. At the individual level, most participants were identified as belonging to the mixed interpretation regime, a small minority of participants as belonging to the probabilistic interpretation of conditionals, and a large minority to the inferentialist interpretation. The interpretation of the fact that inferential considerations were not as clearly dominant in the Negative link condition remains open to speculation. It is reasonable to suppose that participants will have had a sense that there is something wrong with these conditionals and, in an effort to make sense of them nonetheless, also considered the possibility that actually the *negation* of the consequent was meant. But it is impossible to say a priori how such a

kind of interpretative uncertainty may have impacted their ratings. We leave this is as a topic for future research.

4 General discussion

Hypothetical Inferential Theory (HIT) is a theory of conditionals that combines the philosophical theory of inferentialism with hypothetical thinking theory from psychology. According to HIT, the meaning of an indicative conditional "If φ, [then] ψ" centrally involves the strength of the inferential connection between its antecedent, φ, and its consequent, ψ. This connection is relevance-driven and satisficing-bounded, in the sense that the default relevant mental model of the conditional is one in which such a connection is present, and that it need only be subjectively strong (rather than deductively valid). In previous work (Douven et al., 2018, 2020), the corresponding principles were called "the principle of relevant inference" and "the principle of bounded inference," respectively. It was high time to state explicitly HIT's implications for the probabilities of conditionals and to subject them to experimental testing. We also wanted to compare HIT with its main rival, the suppositional account, and to extend our evidence database to include everyday conditionals. Those were the main goals of the two experiments presented in this paper.

In both experiments, participants were given three tasks drawing on the same fixed set of everyday conditionals. One task required participants to judge the probability of the conditionals, a second required them to judge the strength of the inferential connection between antecedent and consequent, and a third asked for conditional probabilities. In Experiment 1, conditional probabilities were measured by asking participants to complete, for each conditional, a probabilistic truth-table task, a method we adopted from Over et al. (2007); in Experiment 2, conditional probabilities were measured by asking participants to engage in suppositional thinking.

According to the principle of relevant inference, the *default* mental representation of conditionals is the one in which there is a strong enough inferential connection between antecedent and consequent. This leaves space for additional, non-default representations, which might be more directly probability-based. Accordingly, we predicted inference strength to be the stronger predictor of the probability of the conditional, with some residual variance explained by conditional probability. This prediction was supported in both experiments.

The principle of relevant inference also does well to explain why in both experiments we found a pattern of individual differences in which most participants were strongly influenced by perceived inference strength, but in which conditional probability judgments still appeared to do some explanatory work for some participants. This minority was small compared to the one found by Skovgaard-Olsen et al. (2019) in their analysis of individual differences. We suspect the difference to be method-driven: the aim of Skovgaard-Olsen and colleagues was to test which normative system was endorsed by participants, and whether this system predicted their response patterns. Thus, their method was based on participants explicitly deciding between two normative explanations provided by fictional characters—inference-based, and Equation-based. Our method, by contrast, did not require explicit acceptance, thus allowing greater margin for implicit processing. This suggests that at least some of the variance explained by conditional probability in the literature might be based on explicit, effortful processing. The pattern fits nicely with HIT's principle of relevant inference, in that conditional probability representations are not default, and hence might require more effortful processing.

Taken together, our findings strongly support HIT, extending the theory to probabilistic evaluations and the evidence to everyday conditionals. Even more importantly, the findings clearly favor HIT over the suppositional account. A possible explanation of the previous empirical support for the latter account (e.g., Evans et al., 2003; Over et al., 2007) is that, in many cases, there is considerable overlap between Equation-based and inference-based probabilities, thus creating a confound. When fully controlling for this confound (as we did), the effect of conditional probability as a predictor becomes marginalized, although it does not entirely vanish.

There are several potential constraints on generality (in the sense of Simons, Shoda, & Lindsay, 2017) on our work. Some of these constraints are a function of running both studies via the INSEAD–Sorbonne University Behavioral Lab. Studies ran via the lab are in French, and most participants are native French speakers aged between 18 and 35. However, they do not have to be students. For the materials, we used everyday conditionals, some of which were taken from the extant literature and translated into French, and some of which we constructed for the current work. We are reasonably confident that our findings generalize to other populations and languages—certainly to UK and USA populations, and English. We are more confident about generalization within WEIRD countries than otherwise, and will in particular welcome attempts to replicate in languages that are not Indo-European, especially those with different linguistic devices for conditionals. We have unpublished findings with a UK population aged 19–75 with similar (but not identical) materials in English, which also show the strong predictive value of inference strength. The wording in that data set was different, and it used a 0–100 scale for inference strength, rather than the Likert scales used in the current study. Although we used Likert-type scales to measure inference strength, we varied the scale points (7 vs. 11, respectively). We thus expect findings to generalize to other scales, including a ratio 1–100 scale. We strongly expect the findings to generalize to other measures of the probability of conditionals, given that we used different measures in the two experiments and got essentially the same outcomes. Lastly, we expect findings to generalize across specific wordings of the questions, and across contents of conditionals, as long as the conditionals refer to everyday events familiar to the participants, and reflect different strengths and categorical types of the inferential link.

We are yet to fully develop a processing model of what we temporarily dubbed in the introduction the *inference* heuristic. At this stage, we think that a dual processing model fits well with extant evidence, but is as yet under-specified: we need more empirical work to narrow it down. The evidence is consistent with a default processing of the inference heuristic, and a more effortful processing of the Equation, but we still need to fill in many gaps. We also need to explore further the relation of the inference heuristic to other heuristics documented in the literature, such as the "*if* heuristic" (Evans & Over, 2004) associated with the suppositional conditional. The *if* heuristic provides the relevance cue that focuses the reasoner on the antecedent of the conditional being true, whereas the *inference* heuristic is linked to the inferential interpretation of the Ramsey Test. In previous work (Douven et al., 2018), we showed that the *if* heuristic can be overridden by providing a strong enough relevance cue, but we still need to map out the relations between these processes.

More generally, we are yet to fully understand the relation between inference-based processes and probability-based processes. Do the processes run in parallel or sequentially? Are they additive or interactive? At this stage we favor a default–interventionist model (Evans, 2007b), that is, one in which the inference heuristic runs first in the sequence, with little effort, sometimes—and later—joined by probabilistic processing; but we have no direct evidence for this. Future work can further explore this hypothesis, for example, by drawing on resource-depletion methods such as concurrent task to load working memory, or rapid responding. Process-tracking meth-

ods, such as latencies, eye-tracking or mouse-tracking, can also provide valuable insights in such future studies (Stewart, Singmann, Haigh, Woods, & Douven, in press). As to the question of whether inference-based and probabilistic processes are independent of each other or interact, our model fitting found some support for interactive models, but we need to explore further the nature of this interaction.[11]

With the more psychological approaches to the suppositional conditional (Evans, 2007a) we share the emphasis on hypothetical thinking and heuristic processes. Admittedly, at this stage, we only offer support for HIT at the computational level, with some preliminary evidence for a processing model. Arguably, though, so did much of the empirical evidence for the Equation (e.g., Evans et al., 2003). While later works reported evidence for processing models, it mainly went to show that full cognitive resources—developmental, individual differences, or training—were required for participants to conform to the Equation (Barrouillet, Gauffroy, & Lecas, 2008; Evans, Handley, Neilens, & Over, 2010; Fugard et al., 2011). This fits effortful processing rather than resource-slim heuristics, and works well with our interpretation of the Equation as non-default processing.

We also need to explore more in depth the role of causality. Particularly relevant is recent work on causal Bayes nets, which generalizes Cheng's (1997) power PC account of causality.[12] HIT appears compatible with this approach as a computational-level model of the inferential relations between antecedent and consequent (Douven et al., 2018). Seeking a tighter integration of inferentialism and HIT with work on causal Bayes nets is an avenue for future research. For recent important work in this direction, see Oaksford and Chater (2020).

The New Paradigm in the psychology of reasoning (Elqayam, 2017; Elqayam & Over, 2013; Oaksford & Chater, 2020) is a family of theories, all sharing a worldview that sees reasoning as a special case of decision-making. It has been characterized as Bayesian (Oaksford & Chater, 2007), although the range of interpretations of this characterization is quite wide (Elqayam & Evans, 2013). *Soft* Bayesians adopt Bayesianism for its psychological advantages—most notably, an acknowledgment of the importance of subjectivity, uncertainty, and degrees of belief in the study of human reasoning—while eschewing the commitment, typical of *strict* Bayesianism, to strong coherence between, on one hand, the probability calculus and Bayesian mechanisms for updating degrees of belief, and, on the other hand, behavior. In this sense, HIT is still within the New Paradigm, albeit firmly on the softer side of Bayesian commitments. HIT fits well with the New Paradigm's focus on knowledge-rich inference (Oaksford & Chater, 2020), although we are yet to develop implications for the social character of reasoning—the second important characteristic of the New Paradigm (Douven, 2021a; Elqayam & Over, 2013; Oaksford & Chater, 2020).

In conclusion, the studies reported here provide an important next step in the direction of developing a more fully-fledged account of how humans process conditionals and of the role that inferential connections play in this. We generalized our previous findings, which focused on abstract conditionals and categorical truth-values, to everyday causal conditionals and probabilistic evaluation. These findings, we argued, are fully compatible with the Ramsey Test in particular

---

[11]As an aside, interaction is what we would expect from a default–interventionist model (Elqayam, 2009).

[12]See Ali, Chater, and Oaksford (2011), Ali, Schlottmann, Shaw, Chater, and Oaksford (2010), Fernbach and Erb (2013), Hall, Ali, Chater, and Oaksford (2016), Oaksford and Chater (2013, 2014, 2017), Skovgaard-Olsen, Stephan, and Waldmann (in press), and van Rooij and Schulz (2019). As Gopnik et al. (2004) note, Cheng's method defines special kinds of Bayes net, known as "noisy-and gates" and "noisy-or gates." For normative reasons to use Cheng's measure for measuring inferential strength, see Douven (2021b).

and the New Paradigm in psychology of reasoning in general. We made further inroads into the quest for a robust psychological theory of conditionals.[13]

References

Adams, E. W. (1975). *The logic of conditionals*. Dordrecht: Reidel.

Adams, E. W. (1998). *A primer of probability logic*. Stanford CA: CSLI Publications.

Alchourrón, C. E., Gärdenfors, P., & Makinson, D. (1985). On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic*, *50*(2), 510–530.

Ali, N., Chater, N., & Oaksford, M. (2011). The mental representation of causal conditional reasoning: Mental models or causal models. *Cognition*, *119*(3), 403–418.

Ali, N., Schlottmann, A., Shaw, A., Chater, N., & Oaksford, M. (2010). Causal discounting and conditional reasoning in children. In M. Oaksford & N. Chater (Eds.), *Cognition and conditionals: Probability and logic in human thinking* (pp. 117–134). Oxford: Oxford University Press.

Andreas, H., & Günther, M. (2019). On the Ramsey test analysis of "because". *Erkenntnis*, *84*(6), 1229–1262.

Andreas, H., & Günther, M. (in press). A Ramsey test analysis of causation for causal models. *British Journal for the Philosophy of Science*. doi: 10.1093/bjps/axy074

Aust, F., Diedenhofen, B., Ullrich, S., & Musch, J. (2013). Seriousness checks are useful to improve data validity in online research. *Behavior Research Methods*, *45*(2), 527–535.

Barrouillet, P., Gauffroy, C., & Lecas, J.-F. (2008). Mental models and the suppositional account of conditionals. *Psychological Review*, *115*(3), 760–771.

Berto, F., & Özgün, A. (in press). Indicative conditionals: Probabilities and relevance. *Philosophical Studies*.

Braine, M. D. (1978). On the relation between the natural logic of reasoning and standard logic. *Psychological Review*, *85*(1), 1–21.

Braine, M. D., & O'Brien, D. P. (1991). A theory of *if*: A lexical entry, reasoning program, and pragmatic principles. *Psychological Review*, *98*(2), 182–203.

Bürkner, P.-C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, *10*(1), 395–411. Retrieved from https://doi.org/10.32614/RJ-2018-017 doi: 10.32614/RJ-2018-017

Bürkner, P.-C., & Charpentier, E. (2020). Modelling monotonic effects of ordinal predictors in bayesian regression models. *British Journal of Mathematical and Statistical Psychology*, *73*(3), 420–451.

Bürkner, P.-C., & Charpentier, E. (2020). Modelling monotonic effects of ordinal predictors in Bayesian regression models. *British Journal of Mathematical and Statistical Psychology*, *73*, 420–451.

Bürkner, P.-C., & Vuorre, M. (2019). Ordinal regression models in psychology: A tutorial. *Advances in Methods and Practices in Psychological Science*, *2*(1), 77-101. Retrieved from https://doi.org/10.1177/2515245918823199 doi: 10.1177/2515245918823199

Carnap, R. (1980). A basic system of inductive logic II. In R. C. Jeffrey (Ed.), *Studies in inductive logic and probability II* (pp. 7–155). Berkeley: University of California Press.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., ... Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software, Articles*, *76*(1), 1–32. Retrieved from https://www.jstatsoft.org/v076/i01 doi: 10.18637/jss.v076.i01

Cheng, P. (1997). From covariation to causation: A causal power theory. *Psychological Review*, *104*, 367–405.

Crupi, V., & Iacona, A. (in press). The evidential conditional. *Erkenntnis*.

Decock, L., & Douven, I. (2011). Similarity after Goodman. *Review of Philosophy and Psychology*, *2*, 61–75.

Delgrande, J. P. (1998). Conditional logics for defeasible reasoning. In D. M. Gabbay & P. Smets (Eds.), *Handbook of defeasible reasoning and uncertainty management systems (vol. 2)* (pp. 135–173). Dordrecht: Kluwer.

de Regt, H. W. (2017). *Understanding scientific understanding*. Oxford: Oxford University Press.

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

Douven, I. (2016a). *The epistemology of indicative conditionals: Formal and empirical approaches*. Cambridge: Cambridge University Press.

Douven, I. (2016b). On de Finetti on nested conditionals. In C. Beierle, G. Brewka, & M. Thimm (Eds.), *Computational models of rationality: Essays dedicated to Gabriele Kern-Isberner on the occasion of her 60th birthday* (pp. 265–279). London: College Publications.

Douven, I. (2017a). How to account for the oddness of missing-link conditionals. *Synthese*, *194*(5), 1541–1554.

Douven, I. (2017b). Inference to the best explanation: What is it? And why should we care? In K. McCain & T. Poston (Eds.), *Best explanations: New essays on inference to the best explanation* (pp. 4–22). Oxford: Oxford University Press.

Douven, I. (2021a). How explanation guides belief change. *Trends in Cognitive Sciences*, *25*, 829–830. doi: 10.1016/j.tics.2021.07.009

Douven, I. (2021b). Tracking confirmation. *Philosophy of Science*, *88*, 398–414.

Douven, I. (in press). *The art of abduction*. Cambridge MA: MIT Press.

Douven, I., Elqayam, S., Gärdenfors, P., & Mirabile, P. (in press). Conceptual spaces and the strength of similarity-based arguments. *Cognition*.

Douven, I., Elqayam, S., & Krzyżanowska, K. H. (in press). Inferentialism: A manifesto. In S. Kaufmann & D. E. Over (Eds.), *Conditionals: Logic, linguistics, and psychology*. London: Palgrave Macmillan.

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., Elqayam, S., Singmann, H., & van Wijnbergen-Huitink, J. (2020). Conditionals and inferential connections: Toward a new semantics. *Thinking & Reasoning*, *26*, 311–351.

Douven, I., & Krzyżanowska, K. H. (2019). The semantics–pragmatics divide: An empirical perspective. In A. Capone, M. Carapezza, & F. L. Piparo (Eds.), *Further advances in pragmatics and philosophy: Theories and applications* (pp. 81–100). Dordrecht: Springer.

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

Douven, I., & Verbrugge, S. (2012). Indicatives, concessives, and evidential support. *Thinking & Reasoning*, *18*(4), 480–499.

Douven, I., & Verbrugge, S. (2013). The probabilities of conditionals revisited. *Cognitive Science*, *37*(4), 711–730.

Douven, I., & Williamson, T. (2006). Generalizing the lottery paradox. *British Journal for the Philosophy of Science*, *57*, 755–779.

Edgington, D. (2014). Conditionals. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*. Retrieved from http://plato.stanford.edu/archives/win2014/entries/conditionals/

Elqayam, S. (2009). Models of dependence and independence: A two-dimensional architecture of dual processing. *Thinking & Reasoning*, *15*(4), 377–387.

Elqayam, S. (2017). The new paradigm in psychology of reasoning. In L. J. Ball & V. A. Thompson (Eds.), *International handbook of thinking and reasoning* (pp. 130–150). Oxford: Routledge.

Elqayam, S., & Evans, J. S. B. T. (2013). Rationality in the new paradigm: Strict versus soft Bayesian approaches. *Thinking & Reasoning*, *19*(3–4), 453–470.

Elqayam, S., & Over, D. E. (2013). New paradigm psychology of reasoning. *Thinking & Reasoning*, *19*(3–4), 249–265.

Evans, J. S. B. T. (2006). The heuristic–analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review*, *13*, 378–395.

Evans, J. S. B. T. (2007a). *Hypothetical thinking: Dual processes in reasoning and judgement*. Hove: Psychology Press.

Evans, J. S. B. T. (2007b). On the resolution of conflict in dual process theories of reasoning. *Thinking & Reasoning*, *13*(4), 321–339.

Evans, J. S. B. T. (2020). The suppositional conditional is not (just) the probability conditional. In S. Elqayam, I. Douven, J. S. B. T. Evans, & N. Cruz (Eds.), *Logic and uncertainty in the human mind: A tribute to David Over* (pp. 57–70). London: Routledge.

Evans, J. S. B. T., Barston, J. L., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition*, *11*, 295–306.

Evans, J. S. B. T., Handley, S. J., Neilens, H., & Over, D. (2010). The influence of cognitive ability and instructional set on causal conditional inference. *Quarterly Journal of Experimental Psychology*, *63*(5), 892–909. Retrieved from https://doi.org/10.1080/17470210903111821 (PMID: 19728225) doi: 10.1080/17470210903111821

Evans, J. S. B. T., Handley, S. J., Neilens, H., & Over, D. E. (2007). Thinking about conditionals: A study of individual differences. *Memory & Cognition*, *35*(7), 1772–1784.

Evans, J. S. B. T., Handley, S. J., & Over, D. E. (2003). Conditionals and conditional probability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(2), 321–335.

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

Fernbach, P. M., & Erb, C. D. (2013). A quantitative causal model theory of conditional reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(5), 1327–1343.

Fugard, A. J. B., Pfeifer, N., Mayerhofer, B., & Kleiter, G. D. (2011). How people interpret conditionals: Shifts toward the conditional event. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(3), 635–648.

Gärdenfors, P. (2000). *Conceptual spaces*. Cambridge MA: MIT Press.

Gauffroy, C., & Barrouillet, P. (2009). Heuristic and analytic processes in mental models for conditionals: An integrative developmental theory. *Developmental Review*, *29*(4), 249–282.

Geis, M. L., & Lycan, W. G. (1993). Nonconditional conditionals. *Philosophical Topics*, *21*, 35–56.

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review*, *111*, 3–32.

Grice, H. P. (1989). Indicative conditionals. In H. P. Grice (Ed.), *Studies in the way of words* (pp. 58–85). Cambridge MA: Harvard University Press.

Hadjichristidis, C., Stevenson, R. J., Over, D. E., Sloman, S. A., Evans, J. S. B. T., & Feeney, A. (2001). On the evaluation of *If p then q* conditionals. In *Proceedings of the twenty-third annual conference of the Cognitive Science Society* (pp. 381–386).

Hall, S., Ali, N., Chater, N., & Oaksford, M. (2016). Discounting and augmentation in causal conditional reasoning: Causal models or shallow encoding? *PLoS ONE*, *11*. Retrieved from e0167741.http://dx.doi.org/10.1371/journal.pone.0167741

Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 201–210). Cambridge: Cambridge University Press.

Kneale, W., & Kneale, M. (1962). *The development of logic*. Oxford: Oxford University Press.

Krzyżanowska, K. H. (2015). *Between "if" and "then"* (Unpublished doctoral dissertation). University of Groningen.

Krzyżanowska, K. H. (2019). What is wrong with false-link conditionals? *Linguistics Vanguard*, *5*, 20190006.

Krzyżanowska, K. H., 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. H., Collins, P. J., & Hahn, U. (2021). True clauses and false connections. *Journal of Memory and Language*, *121*, 104252.

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

Krzyżanowska, K. H., Wenmackers, S., & Douven, I. (2013). Inferential conditionals and evidentiality. *Journal of Logic, Language and Information*, *22*(3), 315–334.

Krzyżanowska, K. H., Wenmackers, S., & Douven, I. (2014). Rethinking Gibbard's riverboat argument. *Studia Logica*, *102*(4), 771–792.

Kyburg, H. E., & Teng, C. M. (2001). *Uncertain inference*. Cambridge: Cambridge University Press.

Liddell, T. M., & Kruschke, J. K. (2018). Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology*, *79*, 328–348.

Mackie, J. L. (1973). *Truth, probability and paradox: Studies in philosophical logic*. Oxford: Oxford University Press.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York: Henry Holt.

McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan*. CRC press.

Mellor, H., & Bradley, R. (in press). Conditionals: Truth, safety, and success. *Mind & Language*, in press.

Mill, J. S. (1843). *A system of logic*. London: Longmans, Green, and Company.

Mirabile, P., & Douven, I. (2020). Abductive conditionals as a test case for inferentialism. *Cognition*, *200*, 104232.

Nelkin, D. K. (2000). The lottery paradox, knowledge, and rationality. *Philosophical Review*, *109*, 373–409.

Oaksford, M., & Chater, N. (2003). Conditional probability and the cognitive science of conditional reasoning. *Mind & Language*, *18*(4), 359–379.

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford: Oxford University Press.

Oaksford, M., & Chater, N. (2010). Causation and conditionals in the cognitive science of human reasoning. *The Open Psychology Journal*, *3*, 105–118.

Oaksford, M., & Chater, N. (2013). Dynamic inference and everyday conditional reasoning in the new paradigm. *Thinking & Reasoning*, *19*(3–4), 346–379.

Oaksford, M., & Chater, N. (2014). Probabilistic single function dual process theory and logic programming as approaches to non-monotonicity in human vs. artificial reasoning. *Thinking & Reasoning*, *20*(2), 269–295.

Oaksford, M., & Chater, N. (2017). Causal models and conditional reasoning. In M. R. Waldmann (Ed.), *The Oxford handbook of causal reasoning* (pp. 327–346). Oxford: Oxford University Press.

Oaksford, M., & Chater, N. (2020). Integrating causal Bayes nets and inferentialism in conditional inference. In S. Elqayam, I. Douven, J. S. B. T. Evans, & N. Cruz (Eds.), *Logic and uncertainty in the human mind* (pp. 116–132). London: Routledge.

Oberauer, K., Weidenfeld, A., & Fischer, K. (2007). What makes us believe a conditional? The roles of covariation and causality. *Thinking & Reasoning*, *13*(4), 340–369.

Over, D. E., & Cruz, N. (in press). Indicative and counterfactual conditionals in the psychology of reasoning. In S. Kaufmann, D. E. Over, & G. Sharma (Eds.), *Conditionals: Logic, linguistics, and psychology*. London: Palgrave Macmillan.

Over, D. E., Douven, I., & Verbrugge, S. (2013). Scope ambiguities and conditionals. *Thinking & Reasoning*, *19*(3–4), 284–307.

Over, D. E., & Evans, J. S. B. T. (2003). The probability of conditionals: The psychological evidence. *Mind & Language*, *18*(4), 340–358.

Over, D. E., Hadjichristidis, C., Evans, J. S. B. T., Handley, S. J., & Sloman, S. A. (2007). The probability of causal conditionals. *Cognitive Psychology*, *54*(1), 62–97.

Paris, J. B., & Vencovská, A. (2017). Combining analogical support in pure inductive logic. *Erkenntnis*, *82*, 401–419.

Pennycook, G., Trippas, D., Handley, S. J., & Thompson, V. A. (2014). Base rates: Both neglected and intuitive. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(2), 544–554.

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

Ramsey, F. P. (1929/1990). General propositions and causality. In D. H. Mellor (Ed.), *Philosophical papers* (pp. 145–163). Cambridge: Cambridge University Press.

Rostworowski, W., Pietrulewicz, N., & Będkowski, M. (in press). Conditionals and specific links: An experimental study. *Synthese*.

Rott, H. (1986). *If*s, *though*, and *because*. *Erkenntnis*, *25*(3), 345–370.

Rott, H. (in press). Difference-making conditionals and the relevant Ramsey test. *Review of Symbolic Logic*, 1–39. doi: 10.1017/s1755020319000674

Simon, H. A. (1982). *Models of bounded rationality*. Cambridge MA: MIT Press.

Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on generality (COG): A proposed addition to all empirical papers. *Perspectives on Psychological Science*, *12*, 1123-1128.

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

Skovgaard-Olsen, N., Kellen, D., Hahn, U., & Klauer, K. C. (2019). Norm conflicts and conditionals. *Psychological Review*, in press.

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

Skovgaard-Olsen, N., Singmann, H., & Klauer, K. C. (2017). Relevance and reason relations. *Cognitive Science*, *41*, 1202–1215.

Skovgaard-Olsen, N., Stephan, S., & Waldmann, M. (in press). Conditionals and the hierarchy of causal queries. *Journal of Experimental Psychology: General*.

Spohn, W. (2015). Conditionals: A unifying ranking-theoretic perspective. *Philosophers' Imprint*, *15*.

Stalnaker, R. (1984). *Inquiry*. Cambridge: Cambridge University Press.

Stewart, A. J., Singmann, H., Haigh, M., Woods, J. S., & Douven, I. (in press). Tracking the eye of the beholder: Is explanation subjective? *Journal of Cognitive Psychology*.

van Fraassen, B. C. (1976). Probabilities of conditionals. In W. Harper & C. Hooker (Eds.), *Foundations of probability theory, statistical inference, and statistical theories of science (vol. 1)* (pp. 261–308). Dordrecht: Reidel.

van Rooij, R., & Schulz, K. (2019). Conditionals, causality and conditional probability. *Journal of Logic, Language and Information*, *28*(1), 55–71.

Vehtari, A., Gabry, J., Magnusson, M., Yao, Y., Bürkner, P., Paananen, T., & Gelman, A. (2020). *loo: Efficient leave-one-out cross-validation and waic for bayesian models.* Retrieved from https://mc-stan.org/loo (R package version 2.3.1)

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432.

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. (2020, July). Rank-normalization, folding, and localization: An improved $\widehat{R}$ for assessing convergence of MCMC. *Bayesian Analysis*. Retrieved from https://doi.org/10.1214%2F20-ba1221 doi: 10.1214/20-ba1221

Woody, A. I. (2004). More telltale signs: What attention to representation reveals about scientific explanation. *Philosophy of Science*, *71*, 780–793.

A   Appendix

A.1   Materials

Note that materials are translated from French.

Causal conditionals used in Experiment 1

If dictatorial regimes take over the West, Western populations will move to the moon.

If bacteria become resistant to antibiotics, doctors will treat headaches with aspirin.

If France builds a machine to control the climate, summers in France will become less hot.

If self-driving cars become more wide-spread, people will not be allowed to drive.

If climate change is curbed, a new ice age will begin.

If a cure for AIDS is discovered, condom sales will drop.

If robots become shy, they will be allowed to vote in general elections.

If France wins the 2020 Eurovision song contest, the 2021 Eurovision song contest will be held in Cannes.

If there is less violence on TV, there will be less violent crime in France.

If more children are sent to boarding school, university applications will decrease.

If there is less traffic on the streets, more children will be able to walk to school.

If more people cycle to work, there will be fewer traffic accidents.

If museums stay open until late, ticket queues will be shorter.

If women get equal pay, there will be more stay-at-home fathers.

If sport facilities in French high schools are improved, France will win more gold medals at the Olympics.

If robotization keeps developing, unemployment rates will increase.

If global warming keeps accelerating, London will be flooded.

If a male pill becomes widely available, there will be fewer abortions.

If European countries maintain the current fishing quotas in the North Sea, many fish species will become endangered.

If it becomes more difficult to get a divorce, fewer people will get married.

If nationalism around the world disappears, there will be fewer conflicts on Earth.

If more people become vegetarian, Apple's stock price will increase.

If the UK organizes a second referendum about exiting the European Union, the majority of voters will vote to stay in the EU.

If the practice of trial by jury is abandoned, more criminals will be declared guilty.

If human genetic engineering becomes possible, the genetically modified human beings will not be allowed to participate in the Olympics.

If classrooms sizes in elementary school are decreased, the literacy rate will increase.

If American forces stay in Syria, there will be more terrorist acts in the US.

If fertility treatments improve, the world population growth will accelerate.

If the global world debt decreases, the world economy will develop.

If there is a plague outbreak in Europe, immigrants will be treated with hostility.

If computers become more powerful, economic predictions will improve.

If genetic research keeps making progress, a cure for cancer will be discovered.

If immigration laws become stricter, the number of immigrants will decrease.

If fruits and vegetables are subsidized, people will have a healthier diet.

If jungle deforestation keeps progressing, gorillas will become extinct.

If the US cuts its oil consumption, global warming will slow down.

If tuition fees are increased, the number of university applications will decrease.

If Sony launches a PlayStation 5, the dividends of the company will increase.

If more people start smoking, more money will be invested in medical research.

If more people wear sunscreen, there will be more cases of skin cancer.

If there are more car owners, there will be more traffic jams.

If nurses' salaries are increased, there will be more nurses.

If Adidas convinces more superstars to wear their sneakers, Adidas shoe sales will increase.

If the price of oil keeps increasing, the price of gas will increase.

If the number of atheists in the world keeps increasing, illegal poaching in Africa will be motivated by economic interests.

If more money is invested in renewable energies, the world population will keep growing.

If taxes on fast food are increased, child obesity will increase.

If parenting is taught in schools, youth criminality will increase.

If more houses are built, there will be more homeless people.

If the US declares Christianity the state religion, more Muslim Americans will convert to Christianity.

Causal conditionals used in Experiment 2

†: Missing link conditionals. *: Negative link conditionals.

---

†If bacteria become resistant to antibiotics, electric scooters will be forbidden in London.

†If robots become shy, Ukraine will be allowed to join the EU.

†If more people become vegetarians, smartphone sizes will keep increasing.

†If the number of atheists in the world keeps increasing, then poaching in Africa will be motivated by economic interests.

†If Brazil wins the next rugby World Cup, the next Nobel Peace Prize will go to a woman.

†If the Notre-Dame Cathedral is repaired by 2021, doctors will cure headaches with aspirin.

†If marijuana use is legalized in Europe, ice melt in the North Pole will be curbed.

†If access to Wikipedia becomes paid, the number of alcoholics will increase.

†If preschools become bilingual, the bedbugs epidemic will be solved.

†If the Pope goes blind, the Great China wall will be invaded by moss.

*If the price of oil keeps increasing, the price of gas will decrease.

*If the tax on fast food is increased, childhood obesity will increase.

*If schools teach parenting skills, youth criminality will increase.

*If more houses are built, the number of homeless people will increase.

*If less violence is shown on TV, there will be more violent crimes in France.

*If sport facilities in French high schools are improved, France will earn fewer gold medals in the Olympic Games.

*If treatments for infertility improve, the growth of the world population will slow down.

*If the US decreases its oil consumption, global warming will speed up.

*If university tuitions are increased, applications to universities will increase.

*If Adidas convinces more superstars to wear their sneakers, Adidas shoe sales will decrease.

If France wins the Eurovision song contest in 2020, the 2021 Eurovision song contest will be held in Cannes.

If more people wear sunscreen, the number of cases of skin cancer will decrease.

If the US declares Christianity the state religion, more Muslim Americans will convert to Christianity.

If France builds a weather control machine, summers in France will become less hot.

If autonomous cars become more widespread, people will not be allowed to drive anymore.

If global warming stops, a new ice age will begin.

If a cure for AIDS is discovered, condom sales will decrease.

If more children to go boarding school, the success rate for the high school diploma will increase.

If there is less traffic on the streets, more children will be able to walk to school.

If more people ride a bike to work, there will be fewer traffic accidents.

If museums stay open until late, ticket queues in museums will be shorter.

If women get equal pay, the number of stay-at-home fathers will increase.

If robotization keeps developing, unemployment rates will increase.

If global warming keeps becoming worse, London will be flooded.

If a male pill becomes largely available, there will be fewer abortions.

If European countries maintain current fishing quotas in the North Sea, many fish species will be endangered.

If it becomes harder to get a divorce, the number of marriages will decrease.

If nationalism across the world disappears, there will be fewer conflicts on earth.

If the use of pesticides is controlled more strictly, the extinction of bees will be avoided.

If class sizes in elementary school are decreased, the national illiteracy rate will decrease.
If US military forces stay in Syria, the number of terrorist acts in the US will increase.
If the world debt decreases, the world economy will improve.
If the plague breaks out in Europe, immigrants will be treated with hostility.
If genetic research keeps making progress, a cure for cancer will be discovered.
If immigration laws become stricter, the number of immigrants will decrease.
If fruits and vegetables are subsidized, people will have a healthier diet.
If jungle deforestation continues, gorillas will go extinct.
If more people start smoking, medical research will receive more money.
If the number of car owners increases, traffic jams will worsen.
If nurses' salaries increase, the number of nurses will increase.
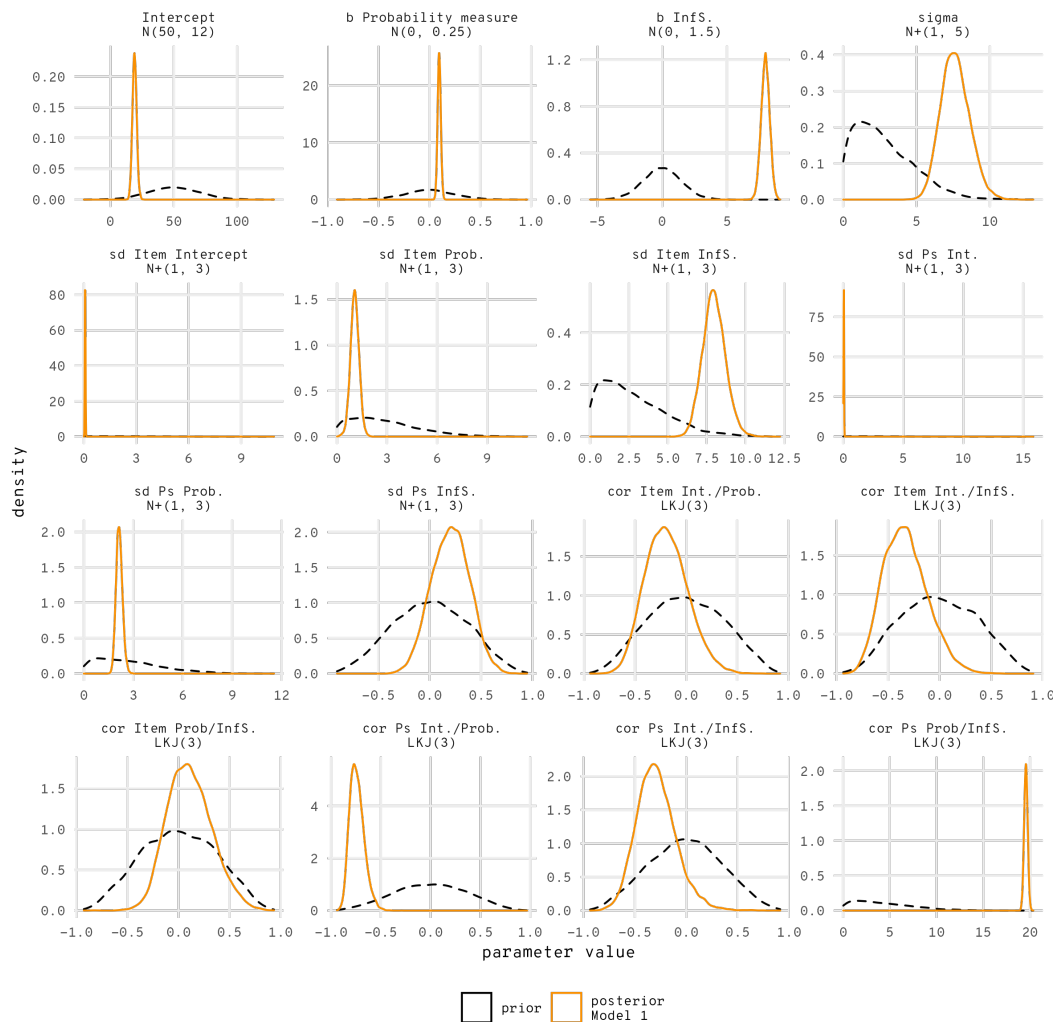
A.2 Prior specifications

Figure 9: Prior distribution of model parameters (black dashed line and facet title) for Model 1 in Experiment 1 and corresponding posterior distribution.
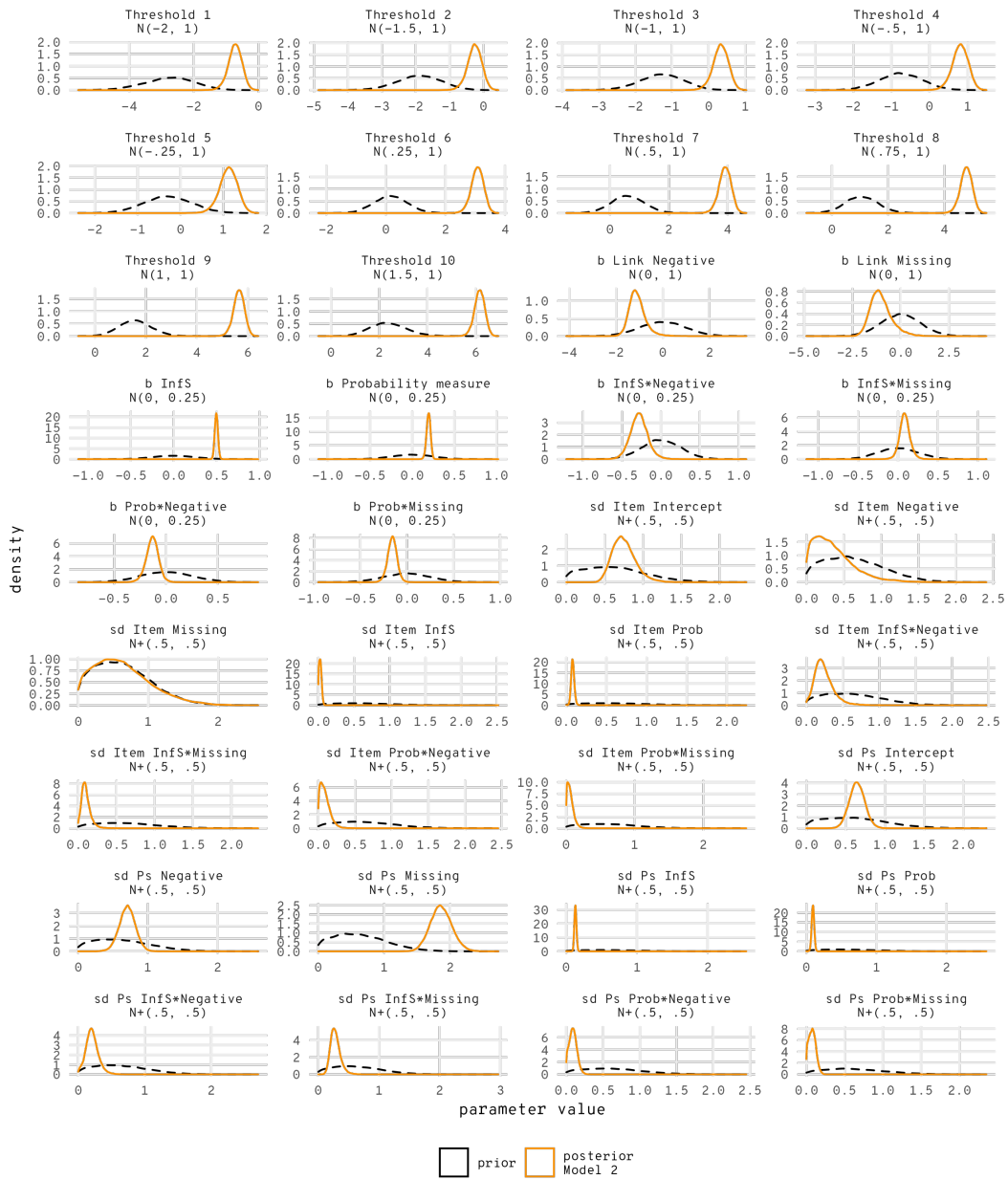
Figure 10: Prior distribution of model parameters (black dashed line and facet title) for Model 2 in Experiment 2 and corresponding posterior distribution.
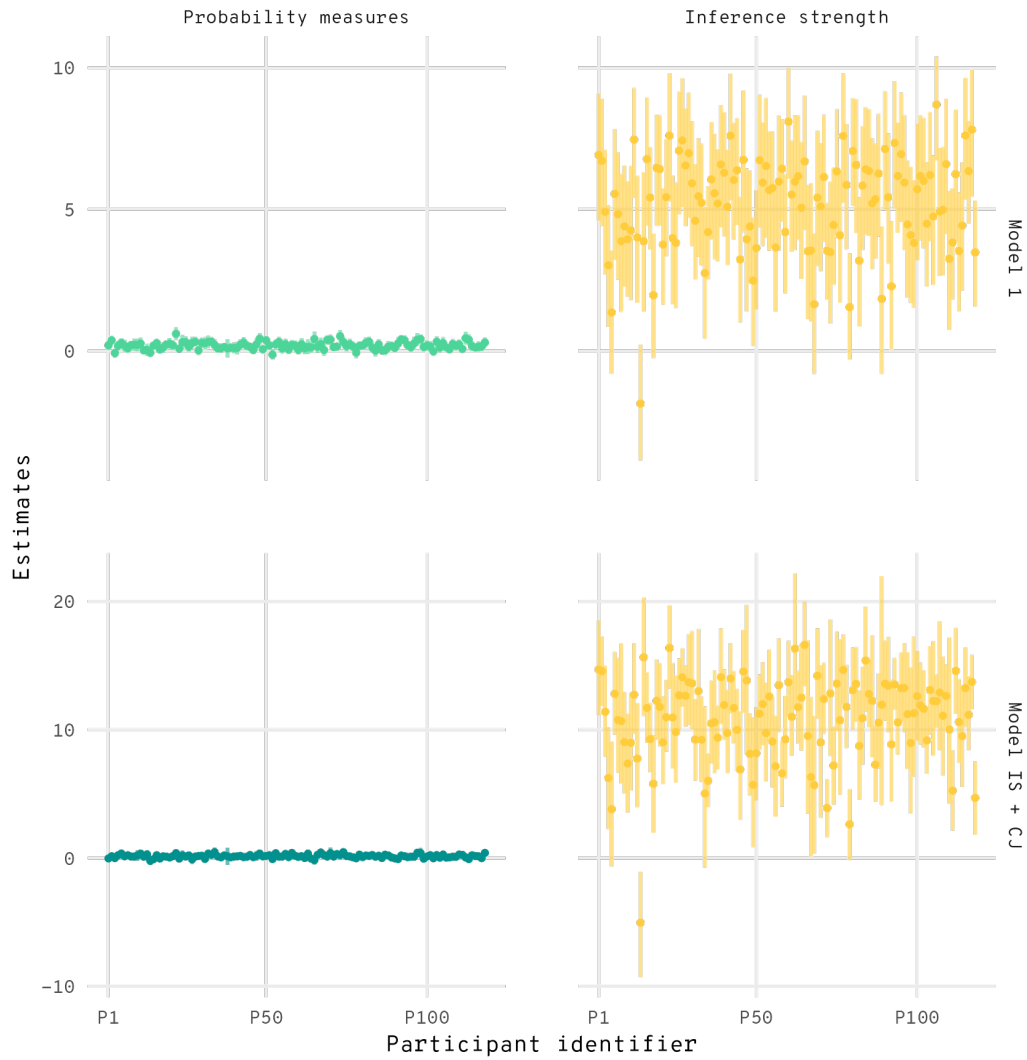
A.3 Individual regressions

Figure 11: Mean posterior probability (with 95 % CI) for the parameter estimates, obtained via individual regressions, for Inference strength and for two probability measures: Conditional probability (Model 1, in light green) and Conjunctive probability (Model IS + CJ, in dark green). (NB: Probabilities were measured on a scale from 0 to 100; inference strength was measured on a 7-point Likert scale.)
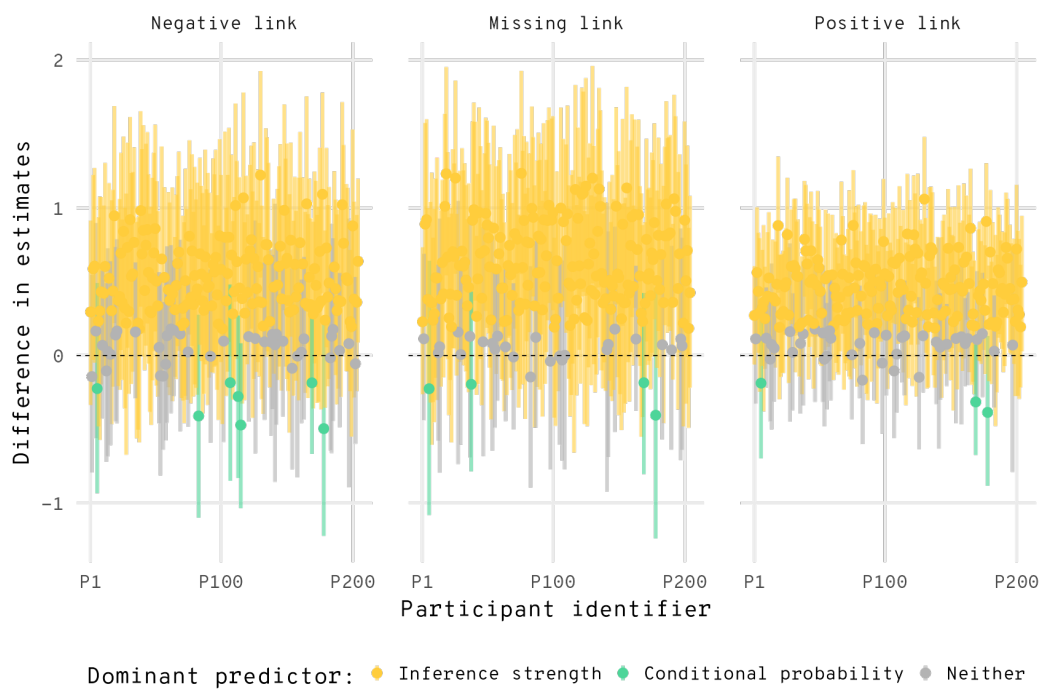
Figure 12: Mean difference in posterior probability (with 95 % CI) between the parameter estimates, obtained via individual regressions, of Inference strength and of Conditional probability (Model 2). A predictor is described as "dominant" when the mean difference between the two predictors is greater than 0.18, or, respectively, smaller than −0.18 (a difference which corresponds, approximately, to a Cohen's *d* of 0.2).