10.1037/xge0000954

Explaining the Implicit Negations Effect in Conditional Inference: Probabilities, Experience, and Contrast Sets

James Vance & Mike Oaksford

Birkbeck College, University of London

Author Note

Department of Psychological Sciences, Birkbeck College, University of London, Malet Street, London, WC1E 7HX, UK. The experiments in this paper were presented at the London Reasoning Workshop, July 2019, at Birkbeck College, University of London. We thank Gernot Kleiter and two anonymous referees for their very helpful comments on this paper. The data for all the experiments are available at https://osf.io/wyjc4/. Please address correspondence to Mike Oaksford (mike.oaksford@bbk.ac.uk).

Abstract

Psychologists are beginning to uncover the rational basis for many of the biases revealed over the last 50 years in deductive and causal reasoning, judgement and decision-making. In this paper, it is argued that a manipulation, experiential learning, shown to be effective in judgement and decision-making may elucidate the rational underpinning of the implicit negation effect in conditional inference. In three experiments, this effect was created and removed by using probabilistically structured contrast sets acquired during a brief learning phase. No other theory of the implicit negations effect predicts these results, which can be modelled using Bayes nets as in causal approaches to category structure. It is also shown how these results relate to a recent development in the psychology of reasoning called "inferentialism." It is concluded that many of the same cognitive mechanisms that underpin causal reasoning, judgement and decision-making may be common to logical reasoning, which may require no special purpose machinery or module.

Keywords: Polarity biases, negations, experiential learning, reasoning biases, new paradigm, causal Bayes nets, inferentialism.

Explaining the Implicit Negations Effect in Conditional Inference: Probabilities, Experience, and Contrast Sets

> "All human systems of communication contain a representation of negation. No animal communication system includes negative utterances, and consequently none possesses a means for assigning truth value, for lying, for irony, or for coping with false or contradictory statements." (Horn, 1989, p. xiii)

The psychology of judgement, decision making, causal, and deductive reasoning reveals many apparent biases. Biases are systematic deviations from the predictions of a normative theory of how people should respond on a task. Explaining these biases is a major industry in cognitive psychology/science that has driven many important theoretical developments. Common patterns of explanation are that the wrong normative theory has been applied to a task (Oaksford & Chater, 1994, 2007; Pothos & Busemeyer, 2013; Pothos, Busemeyer, Shiffrin, & Yearsley, 2017); that people are responding to a different question that has an equally normative answer (Griffths, & Tenenbaum, 2005; Tentori, Crupi, & Russo, 2013); the information was not presented in an understandable format (Gigerenzer & Hoffrage, 1995; Hogarth, & Soyer, 2011; Jarvstad, Hahn, Rushton, & Warren, 2013; Wulff, Mergenthaler-Canseco, & Hertwig, 2018); we need to take account of noise (Costello & Watts, 2014; Costello, Watts, & Fisher, 2018); or that the mind/brain approximates probabilities by sampling (Dasgupta, Schulz, & Gershman, 2017; Hattori, 2016; Sanborn & Chater, 2016; Stewart, Chater, & Brown, 2006), an approach aligned with the classical strategy in the psychology of deductive reasoning of explaining biases at the algorithmic not computational level (Johnson-Laird, 1983; Rips, 1994). Most of these explanations explain away biases while retaining the normative standard of rationality given by classical binary logic (mental logic/mental models) or Bayesian probability theory.[1] That we are beginning to understand the sources of bias in judgement and decision making also resolves a paradox. Explaining biases in the psychology of deductive reasoning, like confirmation bias, has invoked Bayesian probability theory as a normative standard (Oaksford & Chater, 1994, 2007, 2020a). Yet, paradoxically, Bayesian reasoning in judgement and decision-making had seemed equally biased. It also opens up the possibility that the way that biases have been explained away in judgement and decision-making may also apply to the psychology of deductive reasoning.

In this paper, we investigate a key outstanding problem in the psychology of conditional inference, that is, reasoning with *if p then q* in English, where *p* is the antecedent and *q* the consequent. Polarity biases occur when negations (“not”) are varied in conditionals (Evans, 1972, 1998; Evans & Lynch, 1973; Oaksford, 2002; Oaksford & Chater, 1994; Oaksford & Stenning, 1992; Oaksford & Mousakowski, 2004; Schroyens, Schaeken, Fias, & d’Ydewalle, 2000; Schroyens, Schaeken, & d’Ydewalle, 2001; Schroyens, Schaeken, Verschueren, & d’Ydewalle, 2000; Yama, 2001). As our opening quotation from Horn (1989) indicates, negations are a defining feature of human linguistic communication. The Aristotelean foundation of logic, the principle of non-contradiction, cannot be formulated without negations (a

[1] An exception is quantum probability (Pothos & Busemeyer, 2013), which represents a different theory based on quantum logic. It can only be viewed as normative for human reasoning if following its dictates is rational. As for classical probability theory, this question depends on showing that not following its prescriptions leads one to accept bets one is bound to lose, the so-called Dutch book (Vineberg, 2011). Demonstrating this seems to rely on showing that, within a context, quantum probability is equivalent to classical probability theory (Pothos, et al., 2017).

proposition *p* cannot be both true and false, i.e., *not* (*p and not p*)). Negations allow us to deny the claims made by others, setting up contradictions that must be resolved by argumentation (Hahn & Oaksford, 2007; Oaksford & Chater, 2020a). Horn (1989, p. xiii) argued that, "…the absolute symmetry definable between affirmative and negative propositions in logic is not reflected by a comparable symmetry in language structure and language use." It may not be surprising therefore, that, when compared to the standard of formal logic, people's reasoning with negations appears biased.

In the conditional inference paradigm, people may be asked whether they endorse inferences like, *if Johnny does not travel to Manchester* (*not p*) *then he takes the train* (*q*)*, He did not take the train* (*not q*)*, therefore he travelled to Manchester* (*p*). This inference has the form of a logically valid *modus tollens* (MT) argument (formally, *if p then q*, $\neg q$, *therefore*, $\neg p$, where "$\neg$" = not). Illogically, people endorse MT more when it has a negated conclusion (for an *if p then q* conditional) than when it has an affirmative conclusion (for an *if* $\neg p$ *then q* conditional), as in our example (Evans, Clibbens, & Rood, 1996; Evans & Handley, 1999). This phenomenon occurs for all four conditionals in the *negations paradigm*, when negations are systematically varied between the antecedent and consequent (*if p then q*, *if p then* $\neg q$, *if* $\neg p$ *then q*, and *if* $\neg p$ *then* $\neg q$). However, this *negative conclusion bias* is subject to a dramatic effect: it disappears by the simple manipulation of using implicit negations in the categorical premise. For example, denying the consequent of our MT inference by asserting *He travelled by car*, rather than *He did not take the train*.

The implicit negation effect occurs not only for MT but also for the logical fallacies of *denying the antecedent* (DA: *if p, then q*, $\neg p$, therefore $\neg q$) and *affirming the consequent* (AC: *if p, then q*, *q*, therefore *p*), and for the other logically valid inference rule of *modus ponens* (MP: *if* *p, then q*, *p*, therefore *q*). For example, the AC inference on *if not A, then not 2* using an explicit negation, *not 2*, produces 61% endorsements of the conclusion, *not A*. In contrast, using an implicit negation, *7*, causes this to fall to 11% (Evans & Handley, 1999, Expt. 3). Although implicit negations remove negative conclusion bias, they do not lead to logical performance. They reduce conclusion endorsements as much for logically valid inferences (MP, MT) as for logical fallacies (DA, AC).

Explanations of this effect may discriminate between the Bayesian new paradigm approach (Oaksford, 2002; Oaksford, Chater, & Larkin, 2000; Oaksford & Chater, 2003, 2007, 2020a), heuristic approaches (Evans, 1998; Evans et al., 1996; Evans & Handley, 1999), and mental models theory (Johnson-Laird & Byrne, 2002; Khemlani, Orenes, & Johnson-Laird, 2012), but the critical tests have never been conducted.[2] Our experiments attempt to provide these tests. They used probability manipulations shown in decision making to improve participants' understanding of a task and to lead to better fits to the data (Jarvstad et al., 2013; Wulf, et al., 2018). We used short experiential learning phases and asked participants for their subjective estimates of the learned probabilities that we used to predict the results on the inference task. This is the first time that discrete experiential learning has been used to manipulate probabilities in deductive reasoning tasks. We predicted that different acquired

[2] One reason why the critical tests were not conducted may be because the effects were mainly observed for abstract materials, not real world thematic materials (Evans, 1998, 2002). Consequently, it seemed that these biases, although present in the lab, may not generalize to raise concerns about any real world behavior. However, the motivations for both main theories, the matching heuristic (Evans, 1998, 2002) and the contrast set account (Oaksford & Chater, 2007; Oaksford, et al., 2000; Oaksford & Stenning, 1992), came from the pragmatics of negation in natural discourse. Like other illusions created in the lab, perceptual (e.g., the Muller-Lyer illusion) or cognitive, they may still be highly instructive about the normal function of the cognitive system (e.g., the importance of prior experience of a carpentered world).

distributions should be able to create or remove the implicit negation effect in conditional inference. No other theory predicts these effects.

We first briefly introduce the probabilistic Bayesian new paradigm approach to conditional reasoning (for a recent review see, Oaksford & Chater, 2020a). We show how the concept of a *contrast set* (Oaksford 2002; Oaksford & Stenning, 1992) can explain the implicit negations effect, and how it can be created and removed by simple probabilistic manipulations. Testing these predictions requires an effective way of manipulating probabilities. Therefore, we then discuss why using experiential learning may prove a useful method, as in judgement and decision-making (Wulf, et al., 2018). We then introduce our first experiment and derive the specific predictions that we tested.

Probabilities and Contrast Sets

The new Bayesian paradigm in human reasoning is a broad church (Oaksford & Chater, 2020a). However, there are several assumptions common to these approaches. First, the conditional is not a binary truth functional operator, as in the standard logic, that licenses the validity of MP and MT and not of AC and DA. Second, the probability of a conditional is the conditional probability, Pr(*if p then q*) = Pr($q|p$).[3] This assumption is called "the Equation" (Edgington, 1995). Third, probabilities are subjective and relate to individuals' degrees of belief. Finally, conditional probabilities are suppositional and determined by the Ramsey test: suppose *p* is true, add it to your stock of beliefs and read off your degree of belief in *q*.

---

[3] In standard logic, which assumes that propositions are true or false, *if p then q* is false is *p* is true and *q* is false, and true otherwise. Consequently, Pr(*if p then q*) = Pr(*p*, *q*) + Pr(¬*p*, *q*) + Pr(¬*p*, ¬*q*), an assignment that is very rarely observed empirically.

There are a variety of sophisticated probabilistic approaches to conditional inference, for example, probability logic (Cruz, Baratgin, Oaksford, & Over, 2015; Evans, Thompson, & Over, 2015; Pfeifer & Kleiter, 2009; Politzer & Baratgin, 2016; Singmann, Klauer, & Over, 2014), belief revision (Eva & Hartmann, 2018; Oaksford & Chater, 2007, 2010b, 2013), and Bayes nets (Ali, Chater, & Oaksford, 2011; Chater & Oaksford, 2006; Fernbach & Erb, 2013; Oaksford & Chater, 2010b, 2013, 2017). We will discuss these in the sequel. For now, as a first approximation, we assume that the probability of a conclusion of an inference is its conditional probability given the categorical premise calculated over a joint probability distribution (JPD) (Anderson, 1995; Oaksford et al., 2000).[4] We can then derive our predictions by considering two JPDs one without (Table 1) and one with contrast sets (Table 2).

Table 1
Learning a new distribution

Pr_0	q	$\neg q$	Pr_1	q	$\neg q$
p	.3	.1		.3	.1
$\neg p$	.3	.3		.1	.5

Contradictory Negation

Suppose your initial beliefs about Johnny's travelling habits are captured by the JPD Pr_0 in Table 1. In this table, *p* and ¬*p* are contradictories, and are treated with "absolute symmetry" (Horn, 1989, p. xiii). If one of these propositions is true the other is false, but finding out that Johnny did not travel to Manchester conveys nothing about where he may have travelled.

[4] In the General Discussion, we show that both the belief revision and Bayes nets accounts make exactly the same prediction as we derive here. We also identify a problem for the belief revision account that is resolved by treating inference as belief update in Bayes nets.

In Pr_0, you are reasonably confident that *if he travels to Manchester* (*p*), *he takes the train* (*q*). Your degree of belief in the conditional is the relevant conditional probability computed over this JPD, $Pr_0(q|p) = .75$. However, you are maximally uncertain about whether he takes the train or not when he does not travel to Manchester ($Pr_0(q|\neg p) = Pr_0(\neg q|\neg p) = .5$). You also know that just less than half of his journeys are to Manchester ($Pr_0(p) = .4$). Now suppose either that you learn (1) from experience or a reasonably reliable informant.

(1) *If Johnny does not travel to Manchester*, *he does not take the train*.

We assume that the result of learning or hearing (1) from a reliable source, leads you to revise your beliefs about Johnny's travelling habits to the JPD Pr_1 in Table 1, in which $Pr_1(\neg q|\neg p) = Pr_1(\neg p, \neg q)/Pr_1(\neg p) = .5/.6 = .833$.[5] In our experiments, we provide people with relevant experience to revise their beliefs from Pr_0 to Pr_1, where Pr_1 implements manipulations designed to test our account of the implicit negations effect. In the sequel, we fit the model to previous data to estimate people's default prior beliefs, Pr_0.

Suppose you then learn that, on a particular journey, Johnny did not take the train. With what probability should you now believe that he did not go to Manchester? We treat this query as the probabilistic equivalent of an AC inference having learned (1), and with *Johnny did not take the train* as the categorical premise. As we have said, for now, we treat he probability of the conclusion of an inference as the conditional probability of the conclusion given the categorical premise calculated over the JPD Pr_1 in Table 1 (Anderson, 1995; Oaksford et al., 2000). So for AC, $Pr_1(\neg p|\neg q) = Pr_1(\neg p, \neg q)/Pr_1(\neg q) = .5/.6 = .833$. As we will see in the sequel, developing

[5] We use "Pr_0" to "Pr_1" generically in this paper to refer to the JPDs that capture a reasoner's beliefs before, Pr_0, and after, Pr_1, receiving information relevant to changing their beliefs about the conditional premise.

this approach to provide a theory of inference at the computational and algorithmic levels does not alter the predictions we now derive for our experiments using the concept of a contrast set.

Contrary Negation: Contrast Sets

Suppose Peter and Mary are discussing how Johnny travelled to Manchester. Peter says *Johnny travelled to Manchester by car*. As we have seen, Mary can deny Peter's assertion either using an explicit negation, *Johnny did not travel to Manchester by car* or an implicit negation, *Johnny travelled to Manchester by train*. In speech, for the former to make the same point as the latter, the stress must fall on *car*, so that Mary is interpreted to mean that Johnny travelled to Manchester by some other mode of transport (Oaksford, 2002; Oaksford & Stenning, 1992). It is a member of this contrast set (other modes of transport) that Mary can use to implicitly deny Peter's assertion without using a negation.[6]

The philosophical and linguistic depiction of negation as otherness—negated statements make a positive reference to something other than the negated proposition—can be traced back to Plato and to Aristotle's account of contrary negation (Horn, 1989). The variety of ways in which people can use and express negation in natural languages (Horn, 1989) means that identifying contrast sets could not be their sole function. However, they can explain polarity biases (Oaksford, 2002; Oaksford & Stenning, 1992; Oaksford, et al., 2000; Schroyens,

[6] Contrast sets are also highly context sensitive and *ad hoc* (Barsalou, 1983; Oaksford, 2002; Oaksford & Stenning, 1992). They may also depend on category structure that relates to individuals like John (Barsalou, Huttenlocher, & Lamberts, 1998). So, if John's trip originated in Dublin or Peter and Mary are talking about it in Dublin rather than in London, *airplane* might more readily come to mind. Conversational pragmatics, cognitive and deictic context, and intonation, can all cue the ad hoc reference class (modes of transport for conveying people for moderate distances over land or sea) against which various contrast set members that can play the same causal role will be more (car) or less (bike) probable (Oaksford & Stenning, 1992).

Verschueren, Schaeken & d'Ydewalle, 2000), and they may be able explain the implicit negations effect.

Table 2.

A joint probability distribution for implicit negations.

	q_1	q_2	q_3	Total
p_1	0.30 (15)	0.04 (3)	0.06 (2)	0.40 (20)
p_2	0.10 (5)	0.04 (1)	0.02 (2)	0.16 (8)
p_3	0.00 (0)	0.22 (11)	0.22 (11)	0.44 (22)
Total	0.40 (20)	0.30 (15)	0.30 (15)	1.00 (50)

Note. Frequencies of occurrence in the learning trials in Experiment 1 are shown in brackets.

Contrast sets explain this effect by their internal probabilistic structure (Oaksford & Chater, 2007; Oaksford et al., 2000). For example, suppose you know some more details about Johnny's travelling habits. You already know that he usually travels to Manchester by train (see, *Contradictory Negation*). Suppose you also know that he rarely travels to Paris but mostly goes by train (but occasionally by plane or ferry), and that when he travels to Dublin, which he does quite frequently, he only takes the plane or ferry. These facts are captured by the JPD in Table 2, where, p_1 = Manchester, p_2 = Paris, p_3 = Dublin, q_1 = train, q_2 = ferry, q_3 = plane. This table expands Pr_1 in Table 1 to include knowledge of contrast set members. That is, destinations to which Johnny travels other than Manchester and modes of transport that he uses other than the train.

As for Pr_1 in Table 1, knowing the distribution in Table 2 may lead someone to accept (1). On being told *Johnny did not travel to Manchester,* they should then still endorse the conclusion of the MP inference on (1), *he did not take the train*, quite strongly, because in the JPD in Table 2, $\Pr(\neg q|\neg p) = (\Pr(p_2, q_2) + \Pr(p_2, q_3) + \Pr(p_3, q_2) + \Pr(p_3, q_3))/(\Pr(p_2) + \Pr(p_3)) = .5/.6 = .833$. However, if told that *Johnny travelled to Paris*, then the probability that *he did not take the train*, $\Pr(\neg q|p_2) = (\Pr(p_2, q_2) + \Pr(p_2, q_3))/\Pr(p_2) = .06/.16 =.375$, which predicts much lower endorsement of MP. We would expect an implicit negations effect.

All other theories of the implicit negation effect argue that it arises solely from using an implicit negation, regardless of probabilistic structure. However, Table 2 suggests that we should be able remove the effect even when using an implicit negation in the categorical premise. If q_3, *he travelled by plane*, is used to affirm the consequent of (1), $\neg q_1$, then Table 2 does *not* predict an implicit negation effect for AC for this conditional. In this JPD, $\Pr(\neg p|\neg q) = (\Pr(p_2, q_2) + \Pr(p_2, q_3) + \Pr(p_3, q_2) + \Pr(p_3, q_3))/(\Pr(q_2) + \Pr(q_3)) = .833$, and $\Pr(\neg p|q_3) = (\Pr(p_2, q_3) + \Pr(p_3, q_3))/\Pr(q_3) = .24/.30 = .80$. Consequently, whether using an explicit negation (AC-Not) or an implicit negation drawn from the contrast set (AC-Con), people should endorse AC almost equally often. This prediction, that the implicit negations effect depends on probabilistic structure, discriminates the probabilistic contrast set theory from all other theories.

Experience: Manipulating Probabilities

Testing these predictions requires manipulating probabilities. Reasoning researchers have manipulated probabilities in many ways, using pre-tested content (Oaksford, et al., 2000; Oaksford, Chater, & Grainger, 1999), frequency formats (Gigerenzer & Hoffrage, 1995) combined with concrete visualizations (stacks of cards) (Oaksford, et al., 1997, 1999), contingency tables, or “probabilistic truth tables” (Evans, Handley, & Over, 2003; Oberauer & Wlihelm, 2003), as in causal judgement (Ward & Jenkins, 1965), a procedure that has also been reversed so participants construct the contingency table given a conditional (Oaksford & Mousakowski, 2004; Oaksford & Wakefield, 2003; Oberauer, 2006; Over, Hadjichristidis, Evans, Handley, & Sloman, 2007), and sequential tasks where trial frequency reflects the probabilities (Fugard, Pfeifer, Mayrhofer, & Kleiter, 2011; Oaksford & Mousakowski, 2004; Oaksford & Wakefield, 2003), and where learning effects are observed (for critiques, see Jubin & Barrouillet, 2019; Oberauer, Weidenfeld, & Hőrnig, 2004). In these experiments, we used experiential learning of probabilities, which leads to improved performance in judgment and decision-making, and which has not used before in reasoning research.

There is an ongoing debate in judgment and decision-making about the description-experience gap (Hertwig, Barron, Weber, & Erev, 2004). The distinction is between using verbal descriptions of decision options or prospects, and allowing probabilities and utilities to be learned trial-by-trial. One key difference is that people's decision-making seems to be more rational (optimal) with experiential learning, "people are more likely to maximize the experienced mean reward than to maximize the expected value in description" (Wulf et al., 2018, p. 160). Improved performance is also found in probabilistic judgement in general, "even the statistically naïve achieved accurate probabilistic inferences after experiencing sequentially simulated outcomes, and many preferred this presentation format" (Hogarth & Soyer, 2011, p.434). Experiential learning seems to allow people to pick up information about utilities and probabilities more readily than descriptions.[7]

No other theory of the implicit negations effect predicts that learning about probabilistically structured contrast sets should be able to create or remove this effect. As we

[7] We provided a similar motivation, based on natural sampling (Gigerenzer & Hoffrage, 1995; Kleiter, 1994), for using sequential selection tasks (Oaksford & Moussakowski, 2004; Oaksford & Wakefield, 2003).

show in the sequel, all these theories assume that people are attempting to build a mental representation of the logical structure of the premises, which include contradictory logical operators. They are assumed to attempt to draw inferences over these representations using a learned or innate logical competence. Implicit negations are assumed only to disrupt the process of building the appropriate logical representation of the surface linguistic forms of the premises.

However, we need some caution about the extent to which experience based learning leads to performance consistent with normative theories. In probability judgements based on Bayes' theorem, samples from the posterior distribution yield close to normative answers because they are most relevant to the question at hand. That is, for example, what is the posterior probability of a woman having cancer given a positive mammogram? (Hogarth & Soyer, 2011). Samples from the prior distribution, showing very few women have breast cancer, are less relevant and lead to fewer normative responses (Hawkins, Hayes, Donkin, Pasqualino, & Newell, 2015). Moreover, summary descriptions of the posterior sample produce median responses even closer to the normative response (Hawkins et al., 2015).

In conditional inference, the most relevant distribution from which we could provide samples are the conditional probabilities that correspond to people's predicted degree of belief in the conclusion of the inferences MP, DA, AC, and MT (see Table 3 below). However, as for probability judgement, providing such samples is rather too close to giving participants the probabilistically correct answer (Hawkins et al., 2015). Although we wanted to exploit the potential benefits of trial-by-trial sampling, we also wanted to assess people's ability to extrapolate from information that they might experience in the real world. Therefore, we used experiential trial-by-trial learning of the JPD in Table 2, to get participants to revise their default prior beliefs, Pr_0, to a new distribution, Pr_1, which implements the focused manipulations that test our account of the implicit negation effect.

In the sequel, we argue that participants learn a representation like a Bayes net over which they draw inferences just as in causal judgement people are assumed to learn causal strengths from similar learning trials (Ward & Jenkins, 1965). We used a discrete learning task where, using our example, participants observe a series of destination/mode of transport pairs (Anderson & Sheu, 1995; Hattori & Oaksford, 2007). The trial-by-trial approach has been used only once before in studying conditional reasoning (Pollard & Evans, 1983). However, those experiments used a continuous rather than a discrete format (Anderson & Sheu, 1995; Hattori & Oaksford, 2007) that focuses attention on the conditional probabilities like providing samples from these distributions (Oaksford & Chater, 1996). We also assess the extent to which people acquire the appropriate distribution by having them reconstruct the contingency table in Table 2.

Experiment 1: MP Manipulation

There have been no empirical investigations of the probabilistic contrast set account of the implicit negation effect. Our first experiment used a learning phase where participants sample the distribution in Table 2 to revise their beliefs (as in the transition from Pr_0 to Pr_1). The experimental design makes it clear that this sample is from the same population as experienced by an informant who asserts (1) as the major premise of the conditional inferences that participants must then evaluate. Consequently, after the learning phase, participants should be in a similar state of belief as the informant asserting the major premise. Following on from our discussion in *Probabilities and Contrast Sets*, the first hypothesis we tested was:

> *Hypothesis 1*. With contrast sets structured as in Table 2, according to the probabilistic theory, but no other, we should observe an implicit negation effect for MP but not AC. So an interaction is predicted in which MP-Not > MP-Con, AC-Not = AC-Con, MP-Con < AC-Con, and AC-Not = MP-Not.

In this experiment, participants drew inferences before and after the learning phase. We presented single event probability descriptions (e.g., 0.8 or 80%) before the pre-learning inference task. In this phase, we predicted that we would observe the default implicit negations effect, based on the default prior (Pr_0), for these materials. Previous evidence showed an implicit negation effect for this conditional (*if* $\neg p$*, then* $\neg q$) for both MP (MP-Con [44%] < MP-Not [89%]) and AC (AC-Con [11%] < AC-Not [61%]) (Evans & Handley, 1999, Experiment 3). Moreover, in a meta-analysis of previous results, the sample size weighted mean decrease in percentage endorsements for explicit vs implicit negations was 42% for MP, and 57% for AC (Evans & Handley, 1999; Schroyens et al., 2000). Consequently, in this experiment we also tested Hypothesis 2:

> *Hypothesis 2*. In the pre-learning inference task, there will be a greater implicit negation effect for AC than MP.

From our Bayesian perspective, people's default prior probability distribution, Pr_0, causes this effect because it differs from Table 2. Hypothesis 1 suggests that the learning task will overcome this default prior and, in the post learning inference task, reveal an effect for MP but not for AC. We also countenance the possibility that in a novel context, people do not apply informative priors based on prior knowledge but use relatively weak uninformative priors.

In decision making, using participants' subjective estimates of learned probabilities, also provides better fits to the data than objective values (Jarvstad et al, 2013). Consequently, in these experiments, on completing the inference task, we asked participants to perform a probability verification task where they reconstructed the JPD in Table 2. This procedure allowed us to check how well participants had learned this distribution by computing the correlation with the objective values. Splitting participants in to high and low correlation groups will also allow us to see how well the probabilities are learned affects inference. We also used these joint probabilities to calculate the relevant conditional probabilities for each inference. We could then test how well these subjective calculated conditional probabilities predicted inference task performance, which leads to our third hypothesis:

Hypothesis 3. The subjective probability estimates for Table 2, when used to calculate the appropriate conditional probabilities, should be good predictors of the odds of endorsing an inference in the inference task, although how well the JPD is learned might moderate this effect.

We also asked participants to rate their confidence in their inference judgements. In these experiments, we asked participants for a categorical judgement, do you endorse the conclusion or its negation? In much previous (e.g., Oaksford et al, 2000) and recent research (Skovgaard-Olsen, Collins, Krzyżanowska, Hahn, & Klauer, 2019), participants are asked to rate how sure or confident they are in, or the extent they agree with, a conclusion. When rescaled, researchers often treat these ratings as proxies for probabilities in subsequent model fitting exercises. Research in decision-making has shown that confidence moderates the strength of the relation between value and choice (e.g., De Martino, Fleming, Garrett, & Dolan, 2013). We therefore also investigated two further mutually exclusive hypotheses:

Hypothesis 4. Subjective probability will directly predict confidence, or

Hypothesis 4'. Confidence will moderate the strength of the relation between subjective probability and inference.

Analysis Strategy

We analyzed our data using Bayesian statistics (McElreath, 2016; Gelman, Carlin, Stern, Dunson, Vehtari, & Rubin, 2013).

Data analysis. All analyses used Bayesian regression implemented in the **rstanarm** package in R (Goodrich, Gabry, Ali, Brilleman, 2018; R Core Team, 2018). We analyzed continuous dependent variables (computed conditional probabilities and confidence) using the **stan_lmer** function. We analyzed the binary inference data with the **stan_glm** and **stan_glmer** functions with a logit link function depending on whether the experiments introduced additional random variables.

Comparing means. We used the R packages **tidybayes** (Kay, 2019) and **emmeans** (Lenth, 2019), to generate samples for each marginal mean. When comparing means, we assumed a region of practical equivalence (ROPE, Kruschke, 2011) of $0 \pm 0.1 \times SD$ of the differences, and report the proportion of the distribution of differences falling outside the ROPE. This procedure avoids the unrealistic assumption of a point null hypothesis. We report this statistic, where the proportion is p, as "$p \notin$ ROPE".[8] We also computed Cohen's d for each comparison. For all means and differences, we report the 95% highest density interval (HDI) in square brackets.

[8] To be precise, we calculated differences as highest minus lowest mean so that the proportion we report is always the proportion greater than $0.1 \times SD$.

Comparing models. To answer specific research questions, we frequently compare different models of the data. We do not report Bayes factors for these comparisons (or when comparing means), because of the problems for this approach created by non-informative improper priors (see, McElreath, 2016 p. 192; Gelman, et al., 2013, pp. 182-4). We based all model comparisons on expected predictive accuracy (Gelman, et al., 2013: Ch. 7). We compare models using the leave-one-out information criterion (LOOIC), which provides an estimate of the pointwise divergence between the predicted posterior distribution and the data (Vehtari, Gelman, & Gabry, 2017), using the **loo** package in R (Vehtari, Gabry, Yao, Gelman, 2019). We also report Bayesian stacking weights, which are the best fitting weights assigned to the models if they were averaged to best predict the data (Yao, Vehtari, Simpson, & Gelman, 2018).

Data visualization. For categorical predictors, estimated marginal means of a posterior distribution were all plotted using the **afex_plot** function from the **afex** package in R (Singmann, Bolker, Westfall, & Aust, 2019). For continuous predictors, we plotted the data using **sjPlot** (Lüdecke, 2018).

Method

Participants. Participants were recruited via Amazon Mechanical Turk (2017). Sample size was determined both classically (Chow, Shao, & Wang, 2008) and by Bayesian estimation using the **propdiff.mblacc** function from the **SampleSizeProportions** package in R (Joseph, du Berger, & Belisle, 1997). Previous data from Evans and Handley (1999) for the AC inference on *if ¬p then ¬q* was used to estimate effects size. Maintaining a 5% chance of a Type 1 error and a 20% chance of Type 2 error, led to very different required sample sizes; classical: 22 (11 in each group), Bayesian: 244 (122 in each group). One of our key predictions is an interaction, and reliably estimating interactions requires sixteen times more data than main effects (Gelman, 2018). Consequently, recruitment aimed for a sample size of between 250 and 300.

All participants who completed the experiment received a small payment (between US$0.50 and US$1.00). Some responses were excluded because they may have been duplicates, either sharing an MTurk ID or an IP address. After exclusions, the sample size was 272. 52% were female and the sample was aged between 18 and 75 with a median age of 31 years (mean = 34.34, SD = 11.94). English was the first language of 97% of participants.

Design and materials. The experiment was a 6 (Inference and Negation [*InfNeg*]: MP-Not, MP-Con, AC-Not, AC-Con, DA, MT) by 2 (learning phase: Pre, Post) completely within subjects design. MP and AC were presented in both explicit (Not) and implicit (Con) forms. DA and MT were included as filler items in this experiment and as a further check on participants' understanding of Table 2.

The materials concerned the proportion of animals that a vet sees of different species (cats, dogs, rabbits) and colours (black, white, brown). These were varied in accordance with Table 1, with p_1 = cats, p_2 = dogs p_3 = rabbits, q_1 = black, q_2 = brown, q_3 = white. Participants also performed a conditional inference task at two points in the experiment. The conditional or major premise had a negated antecedent and consequent (*if* $\neg p_1$ *then* $\neg q_1$). Participants were told:

"*The vet is considering the following rule about the animals that she sees:*

If it is not a cat, then it is not black.

The vet is told that the next animal she will see is:

One of the following categorical or minor premises was presented for each question: *not a cat* (MP-Not), *a dog* (MP-Con), *not black* (AC-Not), *white* (AC-Con), *a cat* (DA), and *black* (MT). Participants were then asked:

"Please select the option below that best describes what she should conclude about the next animal."

Responses were gathered using a 2AFC procedure with the alternatives determined by the inference:

MP and DA alternatives:	AC and MT alternatives:
That the animal is not black	*That the animal is not a cat*
That the animal is black	*That the animal is a cat*

The alternatives in each pair were presented in random order. According to Table 1, the probability that participants should draw each inference is shown in Table 3.

Table 3

The Probabilities of Drawing Each Inference in Experiment 1

	Negation	
Inf.	Explicit (Not)	Implicit (Con)
MP	.833 (Pr($\neg q_1 \mid \neg p_1$))	.375 (Pr($\neg q_1 \mid p_2$))
AC	.833 (Pr($\neg p_1 \mid \neg q_1$))	.800 (Pr($\neg p_1 \mid q_3$))
DA	.750 (Pr($q_1 \mid p_1$))	
MT	.750 (Pr($p_1 \mid q_1$))	

Note: Inf. = Inference

The experiment also included a learning phase with 50 trials. Each trial consisted of a photograph of one of the 50 animal/colour pairings shown in Table 1. Each photograph showed only the animal against a white background. Each of the 50 photographs was unique. So, for example, participants would see 15 different black cats, and so on. The photographs were cropped and re-sized so that they were the same size and fitted on to a single screen at typical resolution for online presentation. The pictures were presented in random order. To try and ensure that participants attended to the stimuli, on each trial, the participant had to answer two questions with three response options each: *What type of animal is this?* (*Dog*, *Cat*, *Rabbit*), and *What colour best describes this animal?* (*Black*, *White*, *Brown*).

Procedure. This experiment was implemented in **surveygizmo** (www.surveygizmo.com), to which participants were directed from **MTurk** (www.mturk.com). Participants first saw an information screen and had to confirm consent by clicking a check box to proceed. All experiments received ethical approval from the Department of Psychological Sciences, Research Ethics Committee. Participants then provided basic demographic information. This part of the experiment was common to all experiments reported here.

In the first *pre-learning* phase of the experiment participants were provided with the proportion of animals that the vet sees of different species (cats, dogs, rabbits) and colours (black, white, brown) as in the cell entries in Table 2. Participants then carried out the *pre-learning* phase inference task. Each of the six inference questions, including the opening information containing the conditional rule, were presented on a single page in random order. Participants provided a response and then moved a slider bar to indicate their confidence in their response. The slider bar was labelled ‘Not at all confident’ at one end and ‘Completely confident’ at the other. Responses were recorded as a number between 1 and 100. Participants were not able to move to the next page until both responses had been made.

The participants were then given instructions for the learning phase, as in the *Design and Materials* section, where they were told they would see a sample of the animals that the vet sees in the surgery. Participants then performed the *post-learning* phase inference task, this time with no information about the proportion of animals. Finally, participants were presented with a probability verification task to check how accurately they could reconstruct the probability distribution in Table 2. Each participants' subjective conditional probabilities of drawing each inference could then be calculated. This task consisted of nine response boxes in a three by three grid labelled animal type (cat, dog, rabbit) on one axis and colour (black, white, brown) on the other, as in Table 2. Participants were instructed to enter how many of the next 100 animals that the vet would see would be in each category (a similar procedure was used in Oaksford & Wakefield, 2003). If participants attempted to proceed without their responses summing to 100, they were returned to this page with an instruction to make sure their responses did add up to 100 and were provided with the total value they initially entered for guidance.

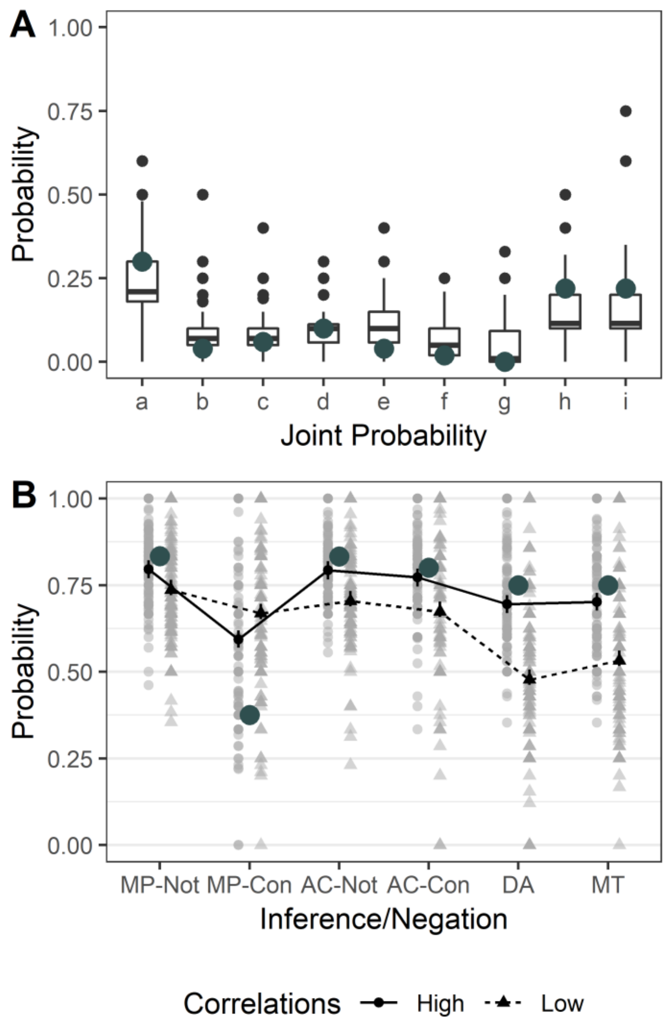
A final page provided participants with a code to enter in MTurk to confirm that they had completed the experiment, thanked them for their time, and provided contact details if they had any questions.

Results and Discussion

Attention test. The attention test in the learning task involved naming the animal and colour on each trial. With fifty trials, each participant could make up to 100 errors. The mean error rate was less than 1% (.70, SD = 2.26). Only 37 participants (13.6%) made more than 1 error and out of these only one made more than 8. This participant made 33 errors. We concluded that most participants paid attention to the stimuli in the learning task and it was not necessary to exclude any participant from subsequent analyses.

Figure 1

Joint Probabilities and Calculated Conditional Probabilities from the Probability Verification Task in Experiment 1



*Note. A. Box-plots for the verification judgements for all cells of Table 1. B. Mean calculated conditional probabilities for each inference based on the estimates shown in panel A split by correlation with the objective values, error bars = 95% HDI; model: Cond ~ InfNeg*Corr . In both panels, the large dark grey points indicate the objective probabilities based on Table 1.*

Probability verification task. We first report the results of the probability verification task. Figure 1A shows the box-plots for each cell in Table 2 and the objective values for each cell. We used the standard letter labelling of cells in a contingency table used in causal learning

(Hattori & Oaksford, 2007). Errors for low probabilities can only push in one direction and all cell values must sum to 1. Therefore, unsurprisingly, lower objective values tended to be overestimated and higher values underestimated. The mean correlation between each participant's estimates and the objective values was $r(7) = .59$ (SD = .33). We split participants into high and low correlation groups (*Corr*); high correlation (≥ median): mean $r(7) = .81$ (SD = .11, $N = 148$), and low correlation (< median): mean $r(7) = .32$ (SD = .30, $N = 124$). By this measure, there was a large group of participants who showed a good understanding of the underlying probabilities, but also a group who did not, sometimes showing negative correlations with the objective values.

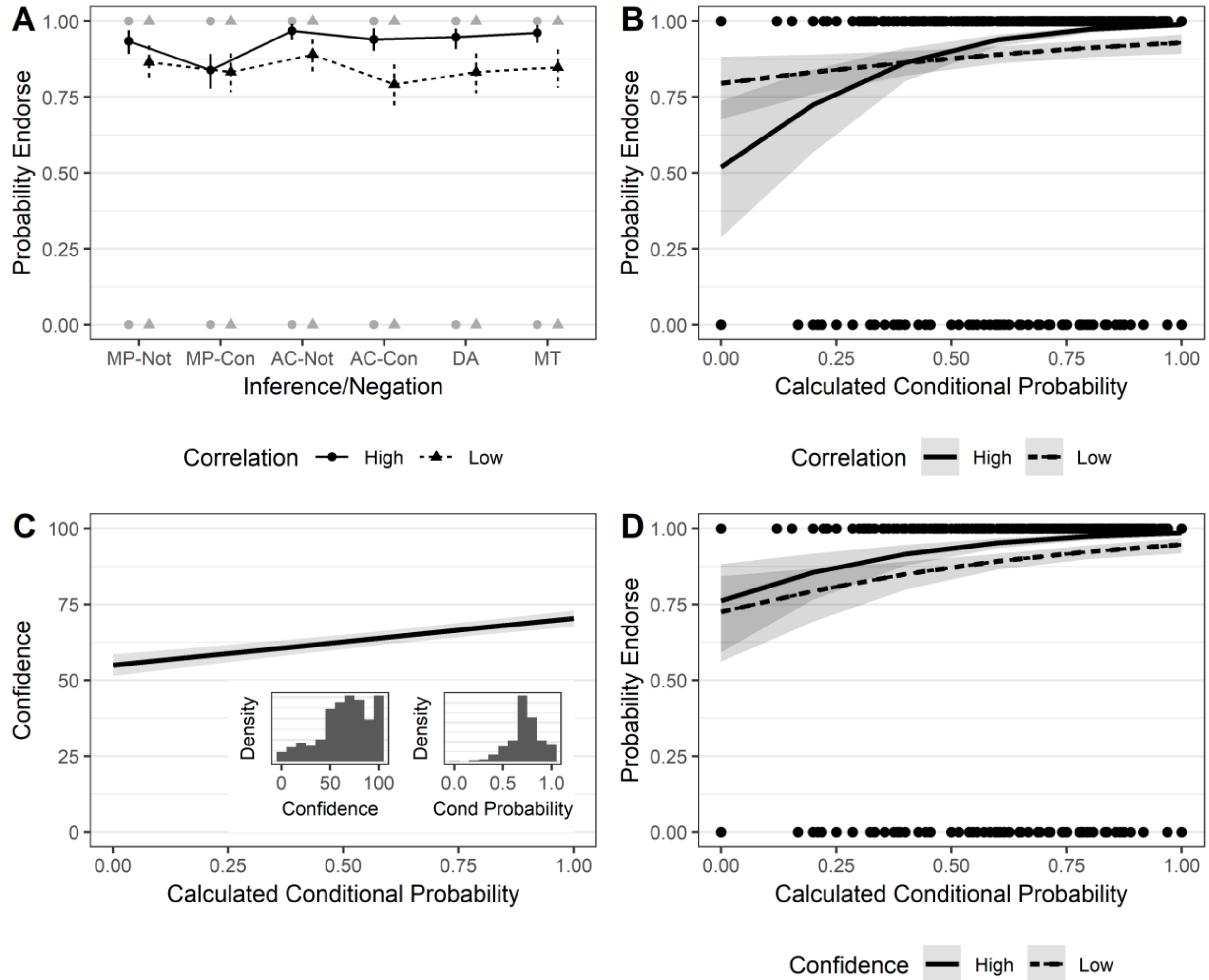
We then used the estimated values from the probability verification task to compute the conditional probabilities (*Cond*) for each inference. There were occasional missing data points because of problems of division by zero. To maintain the coherence of the computed conditional probabilities, rather than impute the missing values, we added .01 to the offending cell(s) in a participants subjective JPD and took .01 from the highest cell value(s). We had to make this adjustment for only 3 participants and 0.49% of cell values and it did not alter the correlations with the objective values. We show the calculated conditional probabilities in Figure 1B, with the data split into high and low correlation groups.

Figure 1B shows the estimated marginal means of the posterior distribution (see figure caption for the model). For the high correlation group, MP-Con (mean = .59, 95% HDI = [.57, .62]) was lower than MP-Not (mean = .80 [.78, .82], $\bar{d}$ = 8.20 [5.88, 10.47], 1.0 ∉ ROPE). Exactly the same pattern of differences was observed between MP-Con and the remaining four inferences, AC-Not (mean = .79 [.77, .82], $\bar{d}$ = 14.90 [12.65, 17.31]), AC-Con (mean = .77 [.75, .79], $\bar{d}$ = 13.64 [11.28, 15.93]), DA (mean = .69 [.67, .72], $\bar{d}$ = 7.69 [5.36, 9.98]), and MT (mean = .70 [.68, .72], $\bar{d}$ = 8.20 [5.88, 10.48]). For all comparisons, 1.0 ∉ ROPE. There were no differences between MP-Not, AC-Not or AC-Con (< .93 ∉ ROPE for all comparisons). For the AC-Not vs AC-Con comparison 0 was a credible value for the effect size, $\bar{d}$ = 1.55 [-.77, 3.77]. However, all these inferences differed from DA and MT (1.0 ∉ ROPE for all comparisons), although DA and MT did not differ from each other (.61 ∉ ROPE). Although the differences were smaller, the same basic pattern occurred for MP-Not, MP-Con, AC-Not, and AC-Con for the low correlation group. However, DA (mean = .48 [.45, .50]) and MT (mean = .53 [.51, .55]) were much lower in the low correlation group than the high correlation group (1.0 ∉ ROPE for both comparisons). In summary, for the high correlation group, the calculated conditional probabilities based on the verification task produced the predicted manipulation such that $\Pr(\neg q_1|\neg p_1)$ (MP-Not) > $\Pr(\neg q_1|p_2)$ (MP-Con), and $\Pr(\neg p_1|\neg q_1)$ (AC-Not) ≈ $\Pr(\neg p_1|q_3)$ (AC-Con).

Inference Tasks. We first looked at the results for the pre-learning inference task with inference (AC, MP) and negation (Not, Con) as categorical predictors. The effect for AC was larger than the effect for MP. AC-Con (mean = .82 [.77, .86]) was lower than AC-Not (mean = .87 [.83, .91]) but zero was still a credible value for the difference ($\bar{d}$ = 2.14 [-.51, 5.06], .94 ∉ ROPE) but only marginally. In contrast, although MP-Con (mean = .83 [.79, .88]) was lower than MP-Not (mean = .86 [.82, .90]) zero was a credible value for the difference ($\bar{d}$ = 1.3 [-1.31, 4.13], .80 ∉ ROPE). No differences were observed between any of the other inferences (0 was a credible value for all differences and < .92 ∉ ROPE for all comparisons). The results of the pre-learning inference task were consistent with default expectations derived from previous research where the implicit negation effect is larger for AC than MP, thereby providing some support for Hypothesis 2. It means that the learning task based on Table 1 must overcome this default prior to reveal the effects predicted by Hypothesis 1.

Figure 2

The Results of the Post-Learning Inference Phase in Experiment 1



Notes. A The probability of endorsing each inference for the high and low correlation groups, error bars = 95% HDI; B. The probability of endorsing an inference predicted by the calculated conditional probability for the high and low correlation groups; C. The relationship between calculated conditional probability and confidence for the high correlation group showing density plots for each variable; D. The probability of endorsing an inference predicted by the calculated conditional probability for the high correlation group with high and low confidence.

We first fitted a model to the post learning phase inference task, using inference/negation and correlation as categorical predictors. The estimated marginal means are shown in Figure 2A. We then looked at the interaction between inference (*Inf*: MP and AC) and negation (*Neg*: Not, Con) for the high correlation group. We compared two models, one which included the interaction (M1), and one with only the main effects (M2) (see, Table 4 Note). *Δelpd* and the Bayesian stacking weights converged on identifying M2 as the best model. It provides the most efficient compression of the data by minimizing the loss of information using the fewest parameters. This result suggests that we have failed to observe the predicted interaction.

However, *Δelpd* indicates that there was only a small difference between models. M2 is weighted more heavily because it is simpler, having fewer parameters. Moreover, estimating interactions requires sixteen times more data than main effects (Gelman, 2018), as we noted in the *Participants* section. The simple effects were as predicted. MP-Con (mean = .84 [.79, .88]) was lower than MP-Not (mean = .93 [.89, .97])]) ($\bar{d}$ = 3.71 [.63, 4.46], .99 $\notin$ ROPE) and AC-Con (mean = .94 [.90, .98]) ($\bar{d}$ = 4.03 [1.22, 6.75], .99 $\notin$ ROPE). However, zero was a credible value for the difference between AC-Not (mean = .97 [.94, .99]) and AC-Con ($\bar{d}$ = 1.57 [-1.30, 4.24], .85 $\notin$ ROPE) and MP-Not ($\bar{d}$ = 1.86 [-.91, 4.75], .89 $\notin$ ROPE).

Table 4

Model Comparison for Predicting Post-Learning Inference Endorsement Rates in Experiment 1

	LOOIC	*SE*	*k*	*ΔLOOIC*	*Δelpd*	*Δse*	*Weight*
M1	324.1	32.3	4.1	1.8	.9	.5	0
M2	322.3	32.0	3.0	0	0	0	1.0

*Note. M1: Endorse ~ Inf*Neg, M2: Endorse ~ Inf + Neg. Estimated number of parameters (k), the difference (ΔLOOIC), the difference in expected log posterior predictive density (Δelpd) and its standard error (Δse), and the Bayesian stacking weights (LOOIC-weight).*

There was only one difference for the low correlation group. AC-Con (mean = .79 [.72, .86]) was lower than AC-Not (mean = .89 [.83, .94]) ($\bar{d}$ = 2.15 [.13, 4.02], .98 $\notin$ ROPE). This effect is consistent with the default prior effect we derived from previous results and the results of the pre-learning inference task. It suggests that even though most participants attended to the learning task, the low correlation group did not learn from it and reverted to the default prior.

The results for the high correlation group confirmed Hypothesis 1. An implicit negation effect can be created (MP) and removed (AC) by varying the underlying probability distribution from which the relevant conditional probabilities are computed. These results are not consistent with other theories of the implicit negations effect.

Calculated conditional probabilities. We next tested whether the calculated conditional probabilities (*Cond*) were good predictors of responses in the inference task (*Endorse*). We also tested whether these probabilities were better predictors of participants’ responses than the logical categorization of the inferences involved. According to other theories, peoples’ responses are driven solely by the logical characterization of the inference involved and whether an explicit or implicit negation is used to express the categorical premise, which is the model we fitted to test Hypothesis 1 (M1). We can compare M1 to a model that uses only the calculated conditional probabilities to predict responses (M3). Fitting this model is equivalent to a repeated measures regression as each participant provides multiple pairs of values (for the current data the six *Cond*/*Endorse* pairs for each level of *InfNeg*) (Bakdash & Marusich, 2017). In hierarchical mixed effects models this is achieved by specifying a different intercept for each participant with the same slope, the population slope (see, Table 5, Note for model specifications). We also fitted a foil model (M4), which included just the intercepts to test that including calculated conditional probability provided more accurate predictions.

Table 5.

Model Comparison for Predicting Endorsement Rates from Calculated Conditional Probabilities in Experiment 1

	LOOIC	*SE*	*k*	*ΔLOOIC*	*Δelpd*	*Δse*	*Weight*
M3	1010.7	50.9	92.5	0	0	0	.89
M4	1038.7	51.8	90.3	-28.0	-14.0	6.0	0
M1	1099.3	52.9	12.6	-88.6	-44.3	11.1	.11

*Note. M3: Endorse ~ Cond*Corr + (1|Participant). M4: Endorse ~ Corr + (1|Participant). Estimated number of parameters (k), the difference (ΔLOOIC), the difference in expected log posterior predictive density (Δelpd) and its standard error (Δse), and the Bayesian stacking weights (LOOIC-weight).*

Table 5 shows the results of the model comparison. The stacking weights and *Δelpd* converged on identifying M3 as the best model. One could argue that M3 provides the better fit because it contains more parameters (*k*). However, Bayesian indices of fit, like LOOIC and BIC, heavily penalize model complexity (many parameters), and far more than conventional fit indices, like AIC[9]. Consequently, that M3 still provides a much better fit is impressive. Moreover, the calculated conditional probabilities are parameter free estimates of the probability

[9] There is a balance to be struck between too many parameters and too few (McElreath, 2016). Too few means important patterns in the data cannot be captured. Too many leads to overfitting, which means that removing data points can lead to large changes in the model's predictions. LOOIC assesses this balance by systematically testing fits by *leaving one out* and ensuring predictions do not radically alter. So that M3 produces the lowest LOOIC value indicates that overfitting is not a problem despite having a greater number of parameters.

of endorsing each inference according to the probabilistic contrast set model. It provides a much better fit because it uniquely predicts the difference between MP-Con and AC-Con. These results confirm Hypothesis 3.

Figure 2B shows the relation between calculated conditional probability and endorsement rates for the high and low correlation groups for M3. Interpreting slopes and interactions is problematic in generalized linear models (Tsai & Gill, 2013). Parameters are estimated after a non-linear logit (i.e., log-odds) transformation of the model. Describing the effects is most interpretable by transforming the dependent variable to odds. The slope for the high correlation group was 129.86 [5.25, 393.63] ($b > 0$, .97 $\notin$ ROPE), that is, a .1 increase in calculated conditional probability increases the odds that an inference will be endorsed by 13. For the low correlation group, the slope was 4.02 [.75, 9.02] ($b > 0$, 1.0 $\notin$ ROPE), that is, a .1 increase in calculated conditional probability increases the odds by .4. The intercept for the high correlation group was 1.29 [.21, 2.94], indicating that when the calculated conditional probability was zero, an inference was still marginally more likely to be endorsed than rejected. For the low correlation group the intercept was 4.74 [.41, 12.28]. Intercepts did not differ between groups ($\bar{d}$ = -1.19 [-.4.32, 1.22], .78 $\notin$ ROPE), but the slope for the high correlation group was steeper than for the low ($\bar{d}$ = 1.11 [.002, 3.44], .95 $\notin$ ROPE).

These results suggest that correlation plays a moderating role. Participants in the high correlation group were more sensitive (lower intercept, higher slope) to changes in the predicted conditional probability when deciding whether to endorse a conclusion than those in the low correlation group. However, there was considerable uncertainty about this relationship for low conditional probabilities. The right hand subplot in Figure 2C shows the density plot for the calculated conditional probabilities. It is skewed towards the upper end of the scale.

Consequently, there were far fewer responses at the lower end explaining the increased uncertainty.

Confidence. We next assessed the relationship between confidence and the calculated conditional probabilities using the model *Confidence ~ Cond + (1|Participant).* Figure 2C shows that they are linearly related. The population slope was 15.38 [10.22, 20.15] ($b > 0$, 1.0 $\notin$ ROPE) indicating that a 0.1 increase in conditional probability lead to 1.54 [1.5, 3.1] point rise in confidence. Both distributions were skewed to the high end of the scale (see subplots in Figure 2C), and they had median values at the same point (conditional probability: .69; confidence: 69). Consistent with this correlation, Figure 2D shows that the median split on confidence (*ConfSplit*) produced a slightly higher intercept when confidence was high without a change in slope (model: _Endorse ~ Cond*ConfSplit + (1|Participant)_). However, zero was a credible value for the differences between high and low response groups for both the slope and the intercept. These results were not consistent with confidence moderating the effect of conditional probability on endorsements. These results, therefore, confirm Hypothesis 4, but disconfirm Hypothesis 4′.

Possible criticisms. Before summarising, we consider two possible criticisms of this experiment. First, the 2AFC response mode may result in more polarized results, perhaps favouring a probabilistic explanation. Response mode can alter response patterns in conditional inference, but not by very much (Evans, Clibbens, & Rood, 1995; Evans & Handley, 1999; Oaksford & Chater, 2010a; Schroyens, Schaeken, & d’Ydewalle, 2001). The 2AFC procedure is similar to evaluation tasks where participants see the valid conclusion and its negation separately and are asked for an endorse decision (Marcus & Rips, 1979; Oaksford, et al., 2000). The current procedure combines these separate choices (which, in the aggregate, sum to 1, see Oaksford, et al., 2000), into a single decision, and provides no reason to expect endorsement decisions to diverge from previously used response modes.

Second, one could argue that in the inference tasks, people are ignoring the conditional premise and are responding solely based on their learned knowledge of the situation. However, one could level this criticism at any attempt to manipulate people's subjective probabilities prior to an inference task in the previous literature. Moreover, the learning phase was short (and were made even shorter in subsequent experiments) and required only that people labelled the items in the attention check, but not learn the probabilistic structure to any criterion of accuracy before proceeding. Finally, of course, this criticism simply begs the question against our Bayesian account, which assumes that to draw inferences people assign relevant conditional probabilities to conditionals based on what they know. They are not applying learned or innate logical rules either syntactically as in mental logic (Rips, 1994), or semantically as in mental models representations (Johnson-Laird, 1983).

Summary. The results of Experiment 1 supported our main hypotheses. Providing single event probabilities for the JPD in Table 2, in the pre-learning phase, led to the standard default effect predicted from previous research confirming H2. There was an implicit negation effect for AC but not MP. In contrast, providing experience of these probabilities, via a brief learning phase, overcame the default priors for the high correlation group consistent with H1. There was an implicit negation effect for MP but not for AC for participants who had learned the JPD. The low correlation group continued to draw inferences consistent with the default prior. The calculated conditional probabilities for each inference, derived from participants' JPD estimates, was also the best predictor of the probability of endorsing an inference (H3). Moreover, confidence was predicted by calculated conditional probability and did not moderate its effect on inference endorsement (H4). These results are not consistent with other theories of the implicit negation effect, which all predict an implicit negation effect for both MP and AC.

Experiment 2: MP and AC Manipulations

Experiment 1 had some limitations. First, the effects, although statistically reliable with good effect sizes, were not of the same magnitude observed in the literature on implicit negations. Moreover, they only occurred for the high correlation group. The low correlation group continued to show the default effect also seen in the pre-learning inference task. Second, although the simple effects were all in the predicted direction, we did not observe the predicted interaction. Third, the distribution of calculated conditional probabilities was skewed toward the upper end of the scale. Such an effect is difficult to avoid when the objective distribution in the JPD (Table 2) were constructed to lead to mainly high conditional probabilities.

Table 6

The distributions of p_i (animals/colours) and q_i (colours/vehicles) used in Experiment 2.

	MP-Manipulation				AC-Manipulation			
	q_1	q_2	q_3	Total	q_1	q_2	q_3	Total
p_1	0.27 (8)	0.00 (0)	0.00 (0)	0.27 (8)	0.27 (8)	0.00 (0)	0.06 (2)	0.33 (10)
p_2	0.06 (2)	0.00 (0)	0.00 (0)	0.06 (2)	0.00 (0)	0.33 (10)	0.00 (0)	0.33 (10)
p_3	0.00 (0)	0.33 (10)	0.33 (10)	0.67 (20)	0.00 (0)	0.33 (10)	0.00 (0)	0.33 (10)
Total	0.33 (10)	0.33 (10)	0.33 (10)	1.00 (30)	0.27 (8)	0.67 (2)	0.06 (20)	1.00 (30)

Note. *p_1 = cats/white, p_2 = dogs/blue, p_3 = rabbits/red, q_1 = black/van, q_2 = brown/car, q_3 = white/motorbike. Frequencies of occurrence in the learning trials using these materials are shown in brackets.*

In Experiment 2, we used a more extreme probability manipulation using the JPDs in Table 6. We also manipulated the JPDs to produce an implicit negation effect for both MP and AC. These changes address all of the limitations of Experiment 1. According to probabilistic contrast set theory a stronger probability manipulation should produce a stronger implicit negation effect. No other theory predicts that this manipulation should have this effect, as they do not make graded predictions. Moreover, by manipulating probabilities in line with the default prior for AC, we should be able to produce a stronger effect, one that may reveal the predicted interaction. By using a more extreme probability manipulation, such that very low calculated conditional probabilities (i.e., zero) are predicted, we may also be able to produce a less skewed distribution, allowing less uncertainty about what is happening at the low end of the scale.

We also reduced the number of learning trials from fifty to thirty. The rationale was part theoretical and part methodological. Theoretically, we have argued that people only build very limited small-scale statistical models related to their immediate deictic or linguistic context (Oakford & Chater, 2020a). These models are constructed on the fly (Chater, 2018) based on linguistic information and prior knowledge, in particular, from immediate past experience, as in decision by sampling models (Stewart, et al., 2006). People's need to predict their immediate environment suggests that they can do so using very few samples (Vul, Goodman, Griffiths, & Tenenbaum, 2014). Methodologically, this experiment used two learning phases. Reducing the number of trials made the experiment more comparable in length to Experiment 1 and less likely to lead to fatigue effects.

We used two sets of materials and participants performed learning phases following by an inference phase for each set of materials in counterbalanced order. We did not use pre-learning inference tasks in this experiment. Consequently, this experiment, and the next, did not evaluate

Hypothesis 2. Participants performed on the MP manipulation for one set of materials and the AC manipulation for the other set of materials. The second set of materials used the colours of motor vehicles and also varied the position of the colour predicates from the consequent to the antecedent clause (see, Table 6), so that the target double negation rule read *if it is not white, then it is not a van*. According to the JPDs in Table 6, the conditional probabilities with which participants should draw each inference for each manipulation are shown in Table 7.

Table 7

The Probabilities of Drawing Each Inference in Experiments 2 and 3

		Negation	
Inf.	Manip.	Explicit (Not)	Implicit (Con)
MP (DA)	MP (DA)	0.91 ($\Pr(\neg q_1 \mid \neg p_1)$)	0.00 ($\Pr(\neg q_1 \mid p_2)$)
	AC (MT)	1.00 ($\Pr(\neg q_1 \mid \neg p_1)$)	1.00 ($\Pr(\neg q_1 \mid p_2)$)
AC (MT)	MP (DA)	1.00 ($\Pr(\neg p_1 \mid \neg q_1)$)	1.00 ($\Pr(\neg p_1 \mid q_3)$)
	AC (MT)	0.91 ($\Pr(\neg p_1 \mid \neg q_1)$)	0.00 ($\Pr(\neg p_1 \mid q_3)$)
DA (MP)	MP (DA)	1.00 ($\Pr(q_1 \mid p_1)$)	
	AC (MT)	0.80 ($\Pr(q_1 \| p_1)$)	
MT (AC)	MP (DA)	0.80 ($\Pr(p_1 \mid q_1)$)	
	AC (MT)	1.00 ($\Pr(p_1 \mid q_1)$)	

Note: *Inf. = Inference; Manip. = Manipulation. The same probability distribution was used in Experiment 3, where it implements the inferences and manipulations shown in parentheses.*

Method

Participants. 334 participants were recruited via **MTurk** after some were excluded because they may have been duplicates or participated in Experiment 1. All participants who completed the experiment received a small payment (between US$0.50 and US$1.00). 53.6% were female and the sample was aged between 18 and 83 with a median age of 36 years (mean = 39.44, SD = 13.32). English was the first language of 96.4% of participants.

Design and Materials. The experiment was a 6 (Inference and Negation: MP-Not, MP-Con, AC-Not, AC-Con, DA, MT) by 2 (Manipulation: MP, AC) completely within subjects design. For each manipulation, participants first carried out a learning task, then the inference task, followed by the probability verification task as in the learning phase of Experiment 1. One set of materials was the same as in Experiment 1. The second set of materials involved vehicles and colours and the new target rule *if it is not white, then it is not a van*. All the relevant substitutions are shown in Table 6 (Note). The order in which participants conducted the task, MP- or AC-manipulation first (Path), and the order of materials, animals or vehicles first (Group), was determined randomly at the beginning of the experiment for each participant. The randomization worked well with roughly equal numbers of participants in the four possible Path by Group conditions (77, 85, 85, 87). Possible artifacts produced by Path or Group were dealt with by treating the four possible Path by Group combinations as a four item random variable (*PaGr*) in mixed effects analyses. In this experiment, the learning phase used only 30 trials.

Procedure. The change from Experiment 1 was that in the two parts of the experiment, participants performed the learning, the inference, and the probability verification tasks in that order. In each part, this procedure was the same as in the learning phase of Experiment 1.

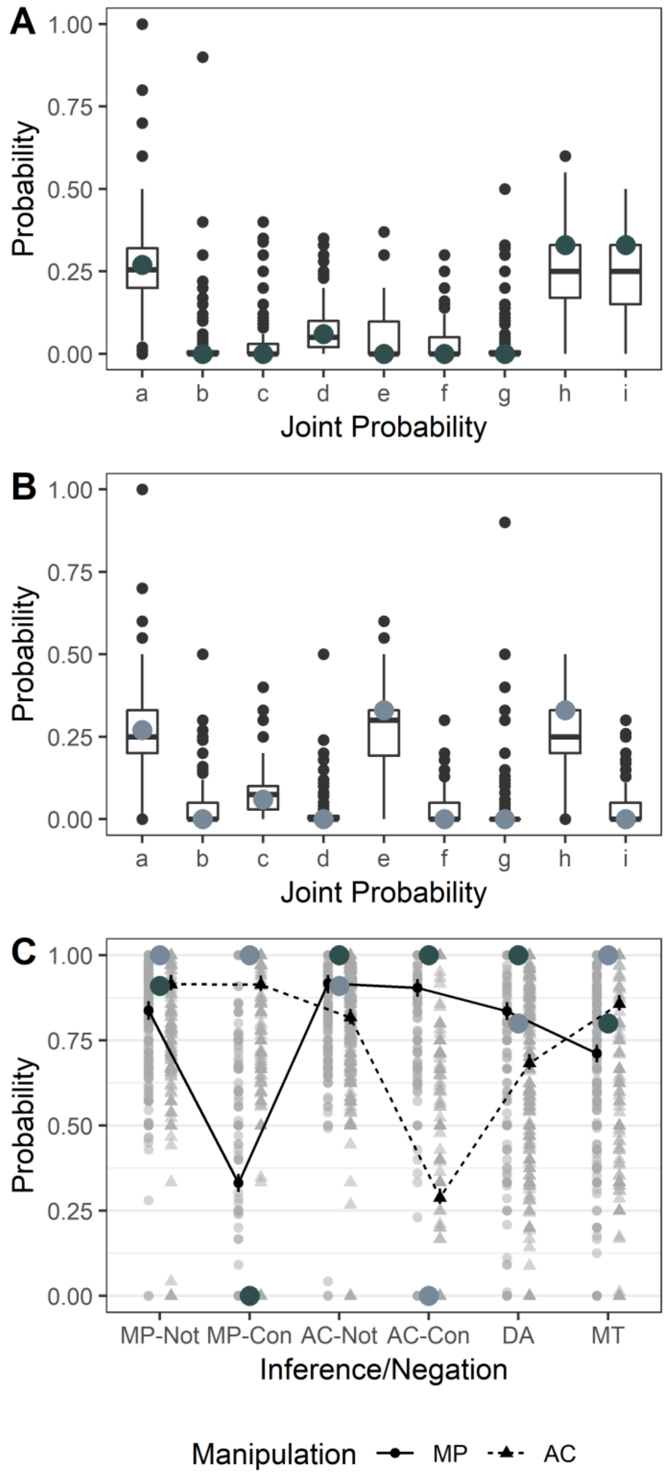
Results and Discussion

Attention test. With two learning tasks with thirty trials in each, each participant could make up to 120 errors. The mean error rate was less than 1% (.80, SD = 4.24). Most participants paid attention to the stimuli in the learning task and no participant was excluded from subsequent analyses.

Probability verification task. Figure 3A and B shows the box-plots for each cell in Table 6 for both the MP- (3A) and the AC-manipulations (3B). The mean correlation between each participant's estimates and the objective values was $r(7) = .74$ (SD = .32). We split participants into high and low correlation groups; high correlation (≥ median): mean $r(7) = .96$ (SD = .04, $N = 167$), and low correlation (< median): mean $r(7) = .52$ (SD = .34, $N = 167$). The average correlations were higher for this cohort than in Experiment 1. If we used the same value for the median as Experiment 1 (.66), then the high group would contain 241 participants and the low group 93. The stronger probability manipulation led to more participants understanding the manipulation. Consequently, we analysed the data without splitting participants in to high and low correlation groups (except when we tested whether the calculated conditional probabilities were good predictors of responses in the inference task).

Figure 3

Joint Probabilities and Calculated Conditional Probabilities from the Probability Verification Task in Experiment 2



Note. A. Box-plots for the verification judgements for all cells of Table 6:MP-Manipulation. B. Box-plots for the verification judgements for all cells of Table 6:AC-Manipulation. C. Mean calculated conditional probabilities for each inference based on the estimates shown in panels A and B, error bars = 95% HDI. In these panels, the large dark grey points indicate the objective probabilities for the MP-Manipulation and the large light grey points indicate the objective probabilities for the AC-Manipulation.

We made the same correction for missing values because of division by zero when calculating conditional probabilities as in Experiment 1, which affected 29 participants (8.7%) and 2.5% of cell values in participants subjective JPDs. Again, this correction did not alter the correlations with the objective values. Figure 3C show the estimated marginal means of the calculated conditional probabilities for each inference split by manipulation (*Manip*). The means were estimated using a linear mixed model, *Cond* ~ *InfNeg***Manip* + (*InfNeg***Manip*|*PaGr*) with the Path and Group variable (*PaGr*) as a random effect to rule out materials and order artifacts.

For the MP-manipulation, MP-Con (mean = .33 [.28, .37]) was lower than MP-Not (mean = .84 [.79, .88]), $\bar{d}$ = 18.67 [16.80, 20.69], 1.0 ∉ ROPE), but zero was a credible value for the difference between AC-Con (mean = .90 [.87, .94]) and AC-Not (mean = .92 [.88, .92]), $\bar{d}$ = .63 [-1.12, 2.32], .70 ∉ ROPE). These results reversed for the AC-manipulation, zero was a credible value for the difference between MP-Con (mean = .91 [.86, .97]) and MP-Not (mean = .91 [.86, .97]), $\bar{d}$ = .03 [-1.81, 1.90], .46 ∉ ROPE), but AC-Con (mean = .29 [.25, .33]) was lower than AC-Not (mean = .82 [.77, .87]), $\bar{d}$ = 19.01 [17.28, 20.59], 1.0 ∉ ROPE). We did not further analyze the results for DA and MT, but note that the calculated conditional probabilities followed the cross over pattern predicted by the objective values. In summary, the calculated conditional probabilities based on the verification task produced the predicted MP-manipulation such that $\Pr(\neg q_1|\neg p_1)$ (MP-Not) > $\Pr(\neg q_1|p_2)$ (MP-Con), and $\Pr(\neg p_1|\neg q_1)$ (AC-Not) ≈ $\Pr(\neg p_1|q_3)$

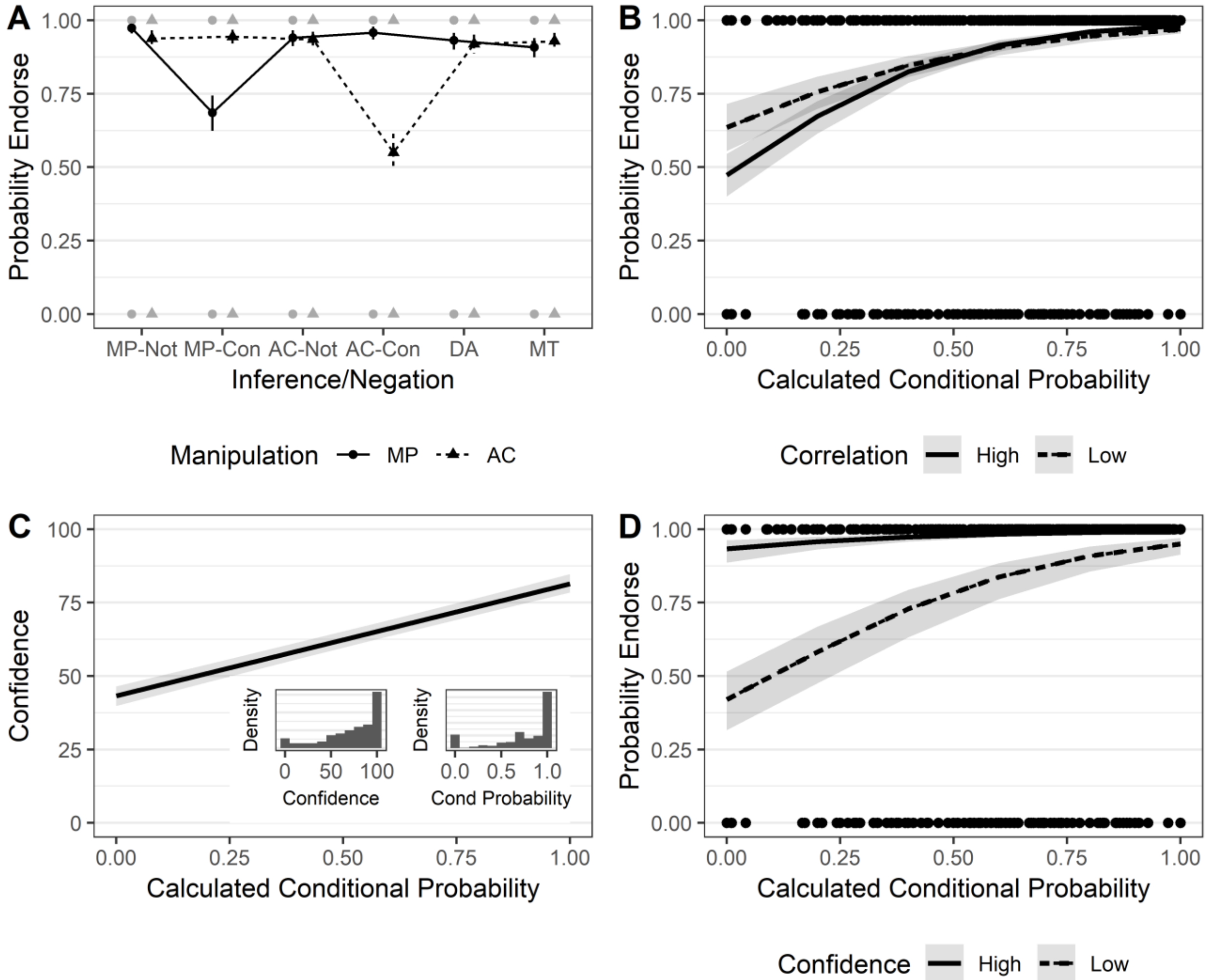
(AC-Con) and the predicted AC-manipulation such that $\Pr(\neg q_1|\neg p_1)$ (MP-Not) ≈ $\Pr(\neg q_1|p_2)$ (MP-Con), and $\Pr(\neg p_1|\neg q_1)$ (AC-Not) > $\Pr(\neg p_1|q_3)$ (AC-Con).

Inference Tasks. We first fitted a model to the inference task, using inference/negation and manipulation as categorical predictors with *PaGr* as a random effect (see, Figure 4A: Notes for the model). We show the estimated marginal means in Figure 4A. We then looked at the interaction between inference (*Inf*: MP and AC) and negation (*Neg*: Not, Con) for each manipulation. As in Experiment 1, we compared two models, one which included the interaction (M1), and one with only the main effects (M2) (see, Table 8: Notes). Table 8 shows the results of the model comparison. The stacking weights and *Δelpd* converged on identifying M1, which includes the interaction, as the best model for both manipulations.

We also assessed the critical simple effects. For the MP-manipulation, the probability of endorsing MP-Con (mean = .68 [.60, .76]) was lower than MP-Not (mean = .97 [.96, .99]), $\bar{d}$ = 7.63 [5.60, 9.57], 1.0 ∉ ROPE), but zero was a credible value for the difference between AC-Con (mean = .96 [.94, .98]) and AC-Not (mean = .94 [.91, .97]), $\bar{d}$ = -1.23 [-3.24, 1.00], .81 ∉ ROPE). These results reversed for the AC-manipulation, zero was a credible value for the difference between MP-Con (mean = .94 [.92, .96]) and MP-Not (mean = .94 [.91, .96]), $\bar{d}$ = -.43 [-2.58, 1.76], .58 ∉ ROPE), but AC-Con (mean = .55 [.50, .60]) was lower than AC-Not (mean = .93 [.91, .96]), $\bar{d}$ = 15.37 [13.23, 17.61], 1.0 ∉ ROPE).

Figure 4

The Results of the Inference Tasks in Experiment 2



*Notes. A. The probability of endorsing each inference for the MP- and AC-manipulations (Endorse ~ InfNeg*Manip + (InfNeg*Manip|PaGr)), error bars = 95% HDI; B. The probability of endorsing an inference predicted by the calculated conditional probability for the high and low correlation groups; C. The relationship between calculated conditional probability and confidence for the high correlation group showing density plots for each variable; D. The probability of endorsing an inference predicted by the calculated conditional probability for the high correlation group with high and low confidence.*

In this experiment, we observed the predicted interactions confirming Hypothesis 1. An implicit negation effect only occurs when the contrast set member used to implicitly negate the antecedent or consequent indicates a low conditional probability of the conclusion. This analysis directly addresses the possible criticism of Experiment 1 that we observed these effects only for the high correlation group. In analyzing these key predictions, in this experiment and the next, we did not split participants by high or low correlation groups.

Table 8

Model Comparison for Predicting Inference Endorsement Rates in Experiment 2

	LOOIC	*SE*	*k*	*ΔLOOIC*	*Δelpd*	*Δse*	*Weight*
MP-Manipulation							
M1	772.8	44.7	8.3	0	0	0	.96
M2	816.4	47.9	7.5	43.6	-21.8	7.1	.04
AC-Manipulation							
M1	934.3	44.8	5.6	0	0	0	.95
M2	971.1	47.0	4.9	36.8	-18.4	6.8	.05

*Notes. M1: Endorse ~ Inf*Neg + (Inf*Neg|PaGr), M2: Endorse ~ Inf + Neg + (Inf + Neg|PaGr). Estimated number of parameters (k), the difference (ΔLOOIC), the difference in expected log posterior predictive density (Δelpd) and its standard error (Δse), and the Bayesian stacking weights (LOOIC-weight).*

Calculated conditional probabilities. We next tested whether the calculated conditional probabilities (*Cond*) were good predictors of responses in the inference task (*Endorse*). We compared the same models as in Experiment 1 but with *PaGr* as a random variable (see Table 9: Notes for the models compared) preserving the maximal random effect structure for each model (Baayen, Davidson, & Bates, 2008). M5 is the model used to generate Figure 4A.

Table 9.

Model Comparison for Predicting Endorsement Rates from Calculated Conditional Probabilities in Experiment 2

	LOOIC	*SE*	*k*	*ΔLOOIC*	*Δelpd*	*Δse*	*Weight*
M3	2170.3	75.7	142.8	0	0	0	.78
M5	2451.1	80.2	16.4	280.8	-140.4	24.7	.22
M4	2751.2	81.9	137.2	580.9	-290.5	26.8	0

*Notes. M3: Endorse ~ Cond*Corr + (1|Participant) + (Cond*Corr|PaGr), M4: Endorse ~ Corr + (1|Participant) + (Corr|PaGr), M5: Endorse ~ InfNeg*Manip + (InfNeg*Manip|PaGr). Estimated number of parameters (k), the difference in LOOICs (ΔLOOIC), the difference in expected log posterior predictive density (Δelpd) and its standard error (Δse), and the Bayesian stacking weights (LOOIC-weight).*

Table 9 shows the results of the model comparison. The stacking weights and *Δelpd* converged on identifying M3 as the best model, confirming the results of Experiment 1 that most information relevant to drawing these inferences is in the predicted conditional probabilities. Figure 4B shows the relation between calculated conditional probability and endorsement rates for the high and low correlation groups for M3. The slope for the high correlation group was 65.57 [34.88, 100.81] ($b > 0$, 1.0 ∉ ROPE), that is, a .1 increase in calculated conditional probability increases the odds that an inference will be endorsed by 6.60. For the low correlation group, the slope was 18.56 [8.88, 30.02] ($b > 0$, 1.0 ∉ ROPE), that is, a .1 increase in calculated conditional probability increases the odds by 1.9. The intercept for the high correlation group was .92 [.59, 1.26], indicating that when the calculated conditional probability was zero, an inference was marginally more likely to be rejected than endorsed. For the low correlation group the intercept was 2.02 [1.10, 3.16]. The intercept was higher for the low correlation group than for the high ($\bar{d}$ = -2.67 [-6.01, .01], .97 ∉ ROPE), and the slope was steeper for the high correlation group than for the low ($\bar{d}$ = 3.57 [1.27, 6.30], 1.0 ∉ ROPE).

Replicating Experiment 1, calculated conditional probability was the best predictor of inference endorsement. This experiment also confirmed that correlation had a moderating effect. With the stronger probability manipulation, better understanding of the probability distribution (high correlation) led to greater sensitivity (a lower intercept and higher slope). The stronger probability manipulation also led to reduced uncertainty at the lower end of the scale, revealing that the intercepts also differed.

Confidence. We next assessed the relationship between confidence and the predicted conditional probabilities. Figure 4C shows that they are linearly related, which we again assessed with separate intercepts for each participant and *PaGr* as a random effect. The population slope was 38.33 [33.44, 43.39] ($b > 0$, 1.0 ∉ ROPE) indicating that a 0.1 increase in conditional probability lead to 3.83 point rise in confidence. Both distributions were skewed to the high end of the scale (see subplots in Figure 4C), and their median values were .89 (conditional probability) and 81 (confidence). Figure 4D shows that in Experiment 2, confidence did not moderate the effect of conditional probability on inference endorsement. Figure 4D is explained by the high correlation between confidence and calculated conditional probability (Figure 4C). Because of this correlation, most of the high calculated conditional probabilities were associated with high confidence. In contrast, the low calculated conditional probabilities were associated with low confidence but also, because of the median split (.89), with many high probability responses. Consequently, only low confidence responses had the spread to reveal the sensitivity of endorsement judgements to variation in calculated conditional probability.

Summary. The stronger probability manipulation used in the learning phase of Experiment 2 strongly confirmed Hypothesis 1. There was an implicit negation effect for MP but not for AC for the MP manipulation, and an implicit negation effect for AC but not for MP for the AC manipulation. Not only were the simple effects significant, a model containing the interaction was a more accurate predictor of the data than a model with only the main effects. The calculated conditional probabilities for each inference derived from participants' JPD estimates, were also the best predictor of the probability of endorsing an inference, confirming Hypothesis 3. Moreover, understanding the probability manipulation moderated the effect, with the high correlation group's inference endorsements showing greater sensitivity to calculated conditional probability (lower intercept, higher slope). In contrast, confidence, although highly correlated with calculated conditional probability, confirming Hypothesis 4, did not moderate its effect on inference endorsement. This result is consistent with previous research that treated judgements of confidence as proxies for probabilities. These results are not consistent with other theories of the implicit negations effect, which all predict an implicit negation effect for both MP and AC regardless of the learned probability manipulation used in these experiments.

Experiment 3: MT and DA Manipulation

We have demonstrated that we can produce or eliminate the implicit negation effect by varying the learned probabilistic structure of the relevant contrast sets for MP and AC. In Experiment 3, we attempted to replicate and generalize these findings to the MT and DA inferences. In this experiment, we also used abstract material to show that we can produce the same probabilistic effects for the materials that first demonstrated the implicit negations effect. We used abstract content involving shapes and colours. The same probability manipulation as in Table 6 achieves the desired result using the conditional *if it is white, then it is a van*. The AC-manipulation then generates an MT-manipulation and the MP-manipulation generates a DA-manipulation. We show the probability of drawing each inference in Table 7. In Experiment 3, p_1 = red/white, p_2 = yellow/blue, p_3 = blue/red, q_1 = circle/van, q_2 = triangle/car, and q_3 = square/motorbike.

Method

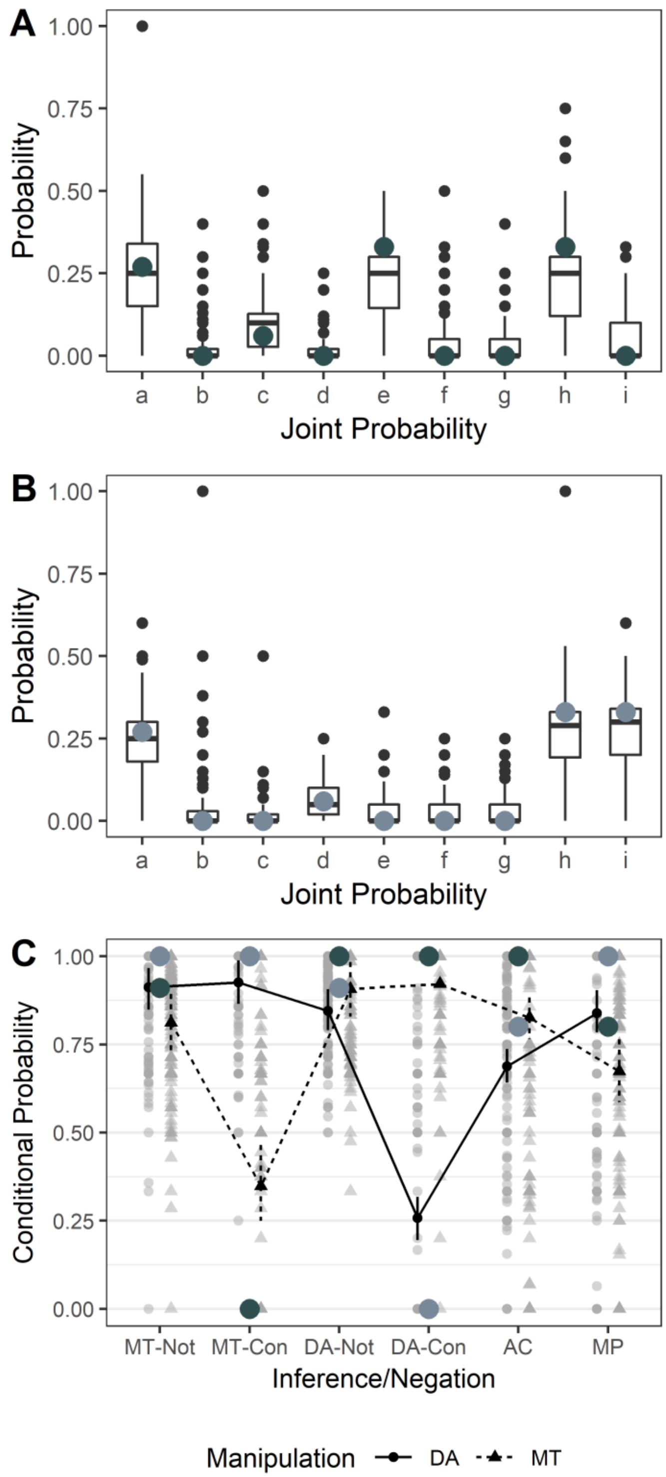
Participants. 168 participants were recruited via **MTurk** after some were excluded because they may have been duplicates or participated in Experiments 1 or 2. All participants who completed the experiment received a small payment (between US$0.50 and US$1.00). 56.0% were female and the sample was aged between 19 and 75 with a median age of 34 years (mean = 38.05, SD = 13.75). English was the first language of 96.4% of participants.

Design and Materials. The experiment was a 6 (Inference and Negation: MT-Not, MT-Con, DA-Not, DA-Con, AC, MP) by 2 (Manipulation: MT, DA) completely within subjects design. One set of materials was the same as in Experiment 2 but using the new target conditional *if it is white, then it is a van*. The second set of materials involved coloured shapes and the target conditional *if it is red, then it is a circle*. For the abstract materials, participants were provided with a back story involving a quality control manager checking the output of a machine printing cards of different shapes and colours (as in Oaksford et al. 2000: Experiment 1). Other than these changes, the design of Experiment 3 was the same as Experiment 2. The randomization worked well with roughly equal numbers of participants in the four possible Path by Group conditions (35, 37, 45, 51).

Procedure. The procedure was the same as in Experiment 2.

Figure 5

Joint Probabilities and Calculated Conditional Probabilities from the Probability Verification Task in Experiment 3



Note. A. Box-plots for the verification judgements for all cells of Table 6:MT-Manipulation. B. Box-plots for the verification judgements for all cells of Table 6:DA-Manipulation. C. Mean calculated conditional probabilities for each inference based on the estimates shown in panels A and B, error bars = 95% HDI. In these panels, the large dark grey points indicate the objective probabilities for the MT-Manipulation and the large light grey points indicate the objective probabilities for the DA-Manipulation.

Results and Discussion

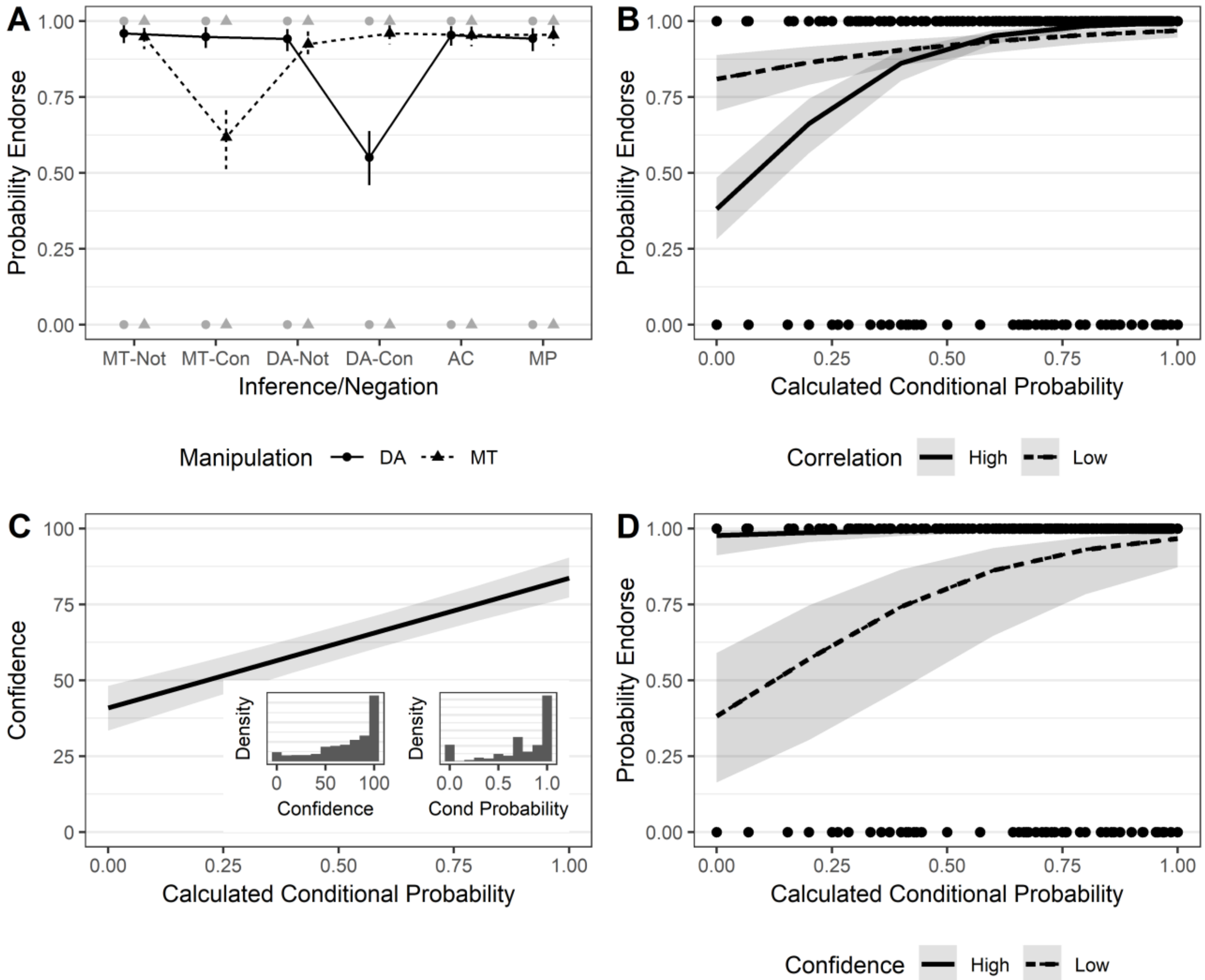
Attention test. The mean error rate (out of 120) was less than 1.0 % (1.10, SD = 4.24). Most participants paid attention to the stimuli in the learning task and so we did not exclude any participants from the subsequent analyses.

Probability verification task. Figure 5A and B shows the box-plots for each cell in Table 5 for both the MT- (5A) and the DA-manipulations (5B). The mean correlation between each participant's estimates and the objective values was $r(7) = .75$ (SD = .32). We split participants into high and low correlation groups; high correlation (≥ median): mean $r(7) = .95$ (SD = .04, $N = 87$), and low correlation (< median): mean $r(7) = .47$ (SD = .34, $N = 81$). As for Experiment 2, we analysed the data without splitting participants in to high and low correlation groups, except when we tested whether the calculated conditional probabilities were good predictors of responses in the inference task.

We made the same correction for missing values because of division by zero when calculating conditional probabilities as in Experiments 1 and 2, which affected 19 participants (11.3%) and 2.4% of cell values in participants subjective JPDs. Again, this correction did not alter the correlations with the objective values. Figure 5C shows the estimated marginal means of the calculated conditional probabilities for each inference split by manipulation (*Manip*). We estimated these means using the same linear mixed model as in Experiment 2.

Figure 6

The Results of the Inference Tasks in Experiment 3



*Notes. A. The probability of endorsing each inference for the MT- and DA-manipulations (Endorse ~ InfNeg*Manip + (InfNeg*Manip|PaGr)), error bars = 95% HDI; B. The probability of endorsing an inference predicted by the calculated conditional probability for the high and low correlation groups; C. The relationship between calculated conditional probability and confidence for the high correlation group showing density plots for each variable; D. The probability of endorsing an inference predicted by the calculated conditional probability for the high correlation group with high and low confidence.*

For the MT-manipulation, MT-Con (mean = .35 [.25, .46]) was lower than MT-Not (mean = .81 [.73, .89]), $\bar{d}$ = 9.62 [7.12, 11.99], 1.0 ∉ ROPE), but zero was a credible value for the difference between DA-Con (mean = .92 [.84, 1.00]) and DA-Not (mean = .91 [.83, .98]), $\bar{d}$

= -.34 [-2.62, 1.92], .59 ∉ ROPE). These results reversed for the DA-manipulation, zero was a credible value for the difference between MT-Con (mean = .93 [.86, .99]) and MT-Not (mean = .91 [.85, .97]), $\bar{d}$ = -.45 [-2.84, 1.80], .62 ∉ ROPE), but DA-Con (mean = .26 [.20, .32]) was lower than DA-Not (mean = .84 [.79, .91]), $\bar{d}$ = 19.47 [17.20, 22.14], 1.0 ∉ ROPE). We did not further analyze the results for AC and MP, but note that the calculated conditional probabilities followed the cross over pattern predicted by the objective values. In summary, the calculated conditional probabilities based on the verification task produced the predicted MT-manipulation such that $Pr(\neg q_1|\neg p_1)$ (MT-Not) > $Pr(\neg q_1|p_2)$ (MT-Con), and $Pr(\neg p_1|\neg q_1)$ (DA-Not) ≈ $Pr(\neg p_1|q_3)$ (DA-Con) and the predicted DA-manipulation such that $Pr(\neg q_1|\neg p_1)$ (MT-Not) ≈ $Pr(\neg q_1|p_2)$ (MT-Con), and $Pr(\neg p_1|\neg q_1)$ (DA-Not) > $Pr(\neg p_1|q_3)$ (DA-Con).

Inference Tasks. We observed no differences for the abstract materials and so we first fitted the same model to the inference task as in Experiment 2 (see, Figure 6A: Notes for the model) with the combined Path and Group variable as a random factor. We show the estimated marginal means in Figure 6A. We then looked at the interaction between inference (*Inf*: MT and DA) and negation (*Neg*: Not, Con) for each manipulation. As in Experiments 1 and 2, we compared s model which included the interaction (M1) with one with only the main effects (M2) (see, Table 10: Notes), and we show the results in Table 10. The stacking weights and *Δelpd* converged on identifying M1, which includes the interaction, as the best model for both manipulations.

We also assessed the critical simple effects. For the MT-manipulation, MT-Con (mean = .62 [.51, .71]) was lower than MT-Not (mean = .95 [.91, .98]), $\bar{d}$ = 8.96 [6.34, 11.57], 1.0 ∉ ROPE), but zero was a credible value for the difference between DA-Con (mean = .96 [.92, .99]) and DA-Not (mean = .92 [.88, .97]), $\bar{d}$ = -1.73 [-4.61, .87], .88 ∉ ROPE). These results reversed for the DA-manipulation, zero was a credible value for the difference between MT-Con (mean = .95 [.91, .98]) and MT-Not (mean = .96 [.93, .99]), $\bar{d}$ = .71 [-2.05, 3.51], .66 ∉ ROPE), but DA-Con (mean = .55 [.50, .60]) was lower than DA-Not (mean = .93 [.91, .96]), $\bar{d}$ = 11.10 [8.34, 13.70], 1.0 ∉ ROPE).

Table 10

Model Comparison for Predicting Inference Endorsement Rates in Experiment 3

	LOOIC	*SE*	*k*	*ΔLOOIC*	*Δelpd*	*Δse*	*Weight*
MT-Manipulation							
M1	453.0	32.4	5.7	0	0	0	.93
M2	478.1	34.2	4.8	25.1	-12.6	5.6	.07
DA-Manipulation							
M1	444.3	32.2	7.1	0	0	0	.86
M2	454.1	33.2	5.9	9.8	-4.9	3.9	.14

*Notes. M1: Endorse ~ Inf*Neg + (Inf*Neg|PaGr), M2: Endorse ~ Inf + Neg + (Inf + Neg|PaGr). Estimated number of parameters (k), the difference (ΔLOOIC), the difference in expected log posterior predictive density (Δelpd) and its standard error (Δse), and the Bayesian stacking weights (LOOIC-weight).*

Replicating Experiment 2, but now for MT and DA, we observed the predicted interactions confirming Hypothesis 1. An implicit negation effect only occurs when the contrast set member used to implicitly negate the antecedent or consequent indicates a low conditional probability of the conclusion.

Calculated conditional probabilities. We next tested whether the calculated conditional probabilities (*Cond*) were good predictors of responses in the inference task (*Endorse*). We compared the same models as in Experiment 2 (see Table 11: Notes for the models compared). M5 is the model used to generate Figure 6A. Table 11 shows the results of the model comparison. The stacking weights and *Δelpd* converged on identifying M3 as the best model, confirming the results of Experiments 1 and 2 that most information relevant to drawing these inferences is in the predicted conditional probabilities. Figure 6B shows the relation between calculated conditional probability and endorsement rates for the high and low correlation groups for M3. The slope for the high correlation group was 365.68 [101.45, 716.17] ($b > 0$, 1.0 $\notin$ ROPE), that is, a .1 increase in calculated conditional probability increases the odds that an inference will be endorsed by 36.5. For the low correlation group, the slope was 8.09 [2.25, 15.30] ($b > 0$, 1.0 $\notin$ ROPE), that is, a .1 increase in calculated conditional probability increases the odds by .81. The intercept for the high correlation group was .64 [.33, 1.00], indicating that when the calculated conditional probability was zero, an inference was marginally more likely to be rejected than endorsed. For the low correlation group the intercept was 7.63 [2.63, 14.48]. The intercept was higher for the low correlation group than for the high ($\bar{d}$ = -2.92 [-5.86, -.76], 1.0 $\notin$ ROPE), and the slope was steeper for the high correlation group than for the low ($\bar{d}$ = 2.64 [.67, 5.21], 1.0 $\notin$ ROPE).

Replicating Experiments 1 and 2, calculated conditional probability was the best predictor of inference endorsement. This experiment also confirmed that correlation had a moderating effect. With the stronger probability manipulation, better understanding of the probability distribution (high correlation) leads to greater sensitivity (lower intercept, steeper slope). Replicating Experiment 2, the stronger probability manipulation led to reduced uncertainty at the lower end of the scale, revealing that the intercepts also differed.

Table 11.

Model Comparison for Predicting Endorsement Rates from Calculated Conditional Probabilities in Experiment 3

	LOOIC	*SE*	*k*	*ΔLOOIC*	*Δelpd*	*Δse*	*Weight*
M3	930.2	50.7	85.9	0	0	0	.85
M5	1173.7	57.5	16.1	243.5	-121.7	21.0	.15
M4	1324.1	58.1	78.8	393.9	-197.0	21.5	0

*Notes. M3: Endorse ~ Cond*Corr + (1|Participant) + (Cond*Corr|PaGr), M4: Endorse ~ Corr + (1|Participant) + (Corr|PaGr), M5: Endorse ~ InfNeg*Manip +* (*InfNeg*Manip|PaGr*)*. Estimated number of parameters (k), the difference in LOOICs (ΔLOOIC), the difference in expected log posterior predictive density (Δelpd) and its standard error (Δse), and the Bayesian stacking weights (LOOIC-weight).*

Confidence. We next assessed the relationship between confidence and the predicted conditional probabilities. Figure 6C shows that they are linearly related, which we again assessed with separate intercepts for each participant and *PaGr* as a random effect. The population slope was 42.88 [30.80, 55.51] ($b > 0$, 1.0 ∉ ROPE), indicating that a 0.1 increase in conditional probability led to a 4.28 point rise in confidence. Both distributions were skewed to the high end of the scale (see subplots in Figure 6C), and their median values were .88 (conditional probability) and 83 (confidence). Figure 6D shows that, replicating Experiment 2, confidence did not moderate the effect of conditional probability on inference endorsement. As for Experiment

2, Figure 6D is explained by the high correlation between confidence and calculated conditional probability (Figure C6).

Summary. Experiment 3 confirmed Hypothesis 1 for MT and DA. There was an implicit negation effect for MT but not for DA for the MT manipulation, and an implicit negation effect for DA but not for MT for the DA manipulation. Not only were the simple effects significant, a model containing the interaction was a more accurate predictor of the data than a model with only the main effects. The calculated conditional probabilities for each inference derived from participants' JPD estimates, were also the best predictor of the probability of endorsing an inference, confirming Hypothesis 3. Moreover, understanding the probability manipulation moderated the effect, with the high correlation group's inference endorsements showing greater sensitivity to calculated conditional probability (lower intercept, higher slope). In contrast, confidence, although highly correlated with calculated conditional probability, confirming Hypothesis 4, did not moderate its effect on inference endorsement. This result is consistent with previous research that treated judgements of confidence as proxies for probabilities. These results are not consistent with other theories, which all predict an implicit negation effect for both MT and DA regardless of the probability manipulation used in these experiments.

General Discussion

Experiments 1 to 3 provided focused experimental tests of the new paradigm probabilistic explanation of the implicit negation effect in conditional inference. We used short discrete learning tasks to impart probabilistic information about contextually limited sets of objects and their properties to manipulate whether an implicitly negated premise would lead to a high or low conditional probability of the conclusion. In Experiment 1, for the high correlation group we observed an implicit negation effect for MP but not for AC, consistent with the probability manipulation. The effects were large in terms of effect size but not of the same apparent magnitude as previously observed. In Experiment 2, we strengthened the probability manipulation and added an AC manipulation to test whether we could elicit and suppress the effect for both inferences. This manipulation produced a much larger effect on calculated conditional probabilities and a correspondingly larger implicit negation effect. We also observed the key interaction showing an implicit negation effect only when predicted by the probability manipulation. Experiment 3 replicated these findings for MT and DA inferences. Across all three experiments, the calculated conditional probability was the best predictor of the odds of endorsing an inference and this effect was moderated by the strength of the correlation between people's judgements of the joint probabilities (Tables 2 and 6) and the objective values. Participants who had better learned the probability distribution (high correlation group) showed greater sensitivity (lower intercept, higher slope) to the calculated conditional probability when endorsing inferences. Calculated conditional probability predicted confidence in whether participants endorsed an inference or not, but confidence did not moderate its effect on inference endorsement. This result is consistent with previous research that used confidence judgements as proxies for probabilities. These results raise a number of issues that we now address. We begin by looking at Bayesian New Paradigm approaches that can implement the predictions that we have just tested.

New Paradigm Probabilistic Approaches

In deriving our predictions we have assumed that the probability of the conclusion of an inference is the conditional probability of the conclusion given the categorical premise.

However, as we indicated in the introduction, this rubric does not provide an account of what people are doing when they learn the categorical premise that provides a theory of inference at either the computational or algorithm level. Fortunately, as we also observed, both approaches we now consider lead to exactly the same predictions that our experiments have just tested.

Belief revision. One approach is to treat inference as belief revision by conditionalization (Eva & Hartmann, 2018; Oaksford & Chater, 2007, 2010b, 2013). This approach provides a computational level theory that justifies our predictions. As we have argued, learning from experience or a reliable informant leads people to revise their degrees of belief from a distribution like Pr_0 to new a distribution like Pr_1 in Table 1. Conditionalization similarly treats learning the categorical premise as belief revision to a new distribution Pr_2. By Jeffrey conditionalization this is achieved via the law of total probability. For example, (2) shows how to calculate the new probability of the conclusion for the MP inference, where you learn a new probability of p, $Pr_2(p)$, that is you come to believe that *Johnny travelled to Manchester* more strongly (> .4).

$$Pr_2(q) = Pr_1(q|p)Pr_2(p) + Pr_1(q|\neg p)Pr_2(\neg p) \qquad (2)$$

If, however, learning p leads to $Pr_2(p) = 1$ (perhaps you think your informant is completely reliable, i.e., Johnny is definitely travelling to Manchester), then (2) reduces to Bayesian conditionalization, where $Pr_2(\neg p) = 0$. Consequently, MP on the conditional *if p then q* in Pr_1 in Table 1 leads to:

$$Pr_2(q) = Pr_1(q|p)Pr_2(p) = Pr_1(q|p) = .75 \qquad (3)$$

That is, the new probability of the conclusion is the old conditional probability of the conclusion given the categorical premise. Consequently, treating inference as Bayesian conditionalization justifies all our predictions.

However, it could be argued that there is a problem with this approach. Take MT on Pr_1 in Table 1, which leads to (4).

$$Pr_2(\neg p) = Pr_1(\neg p|\neg q)Pr_2(\neg q) = Pr_1(\neg p|\neg q) = .833 \quad (4)$$

In the new distribution Pr_2, $Pr_2(q) = 0$, and hence $Pr_2(q|p) = 0$. So in Pr_2, we should no longer find the conditional premise acceptable. That the probability of the conditional premise is not invariant across the belief update means that it is difficult to regard the revision to Pr_2 as capturing what it means to draw these inferences. This set of four logical inferences concern what follows from the premises assumed true or highly probable. Indeed, given (4), this approach seems to imply that we should now believe that Johnny never travels anywhere by train.

However, this argument turns on an equivocation between our enduring beliefs versus how they allow us to draw inferences from the momentary and changing flow of information we experience. Learning about the conditional premise involves adjusting your enduring beliefs about Johnny's travelling habits (the transition from Pr_0 to Pr_1). However, learning the categorical premise in inference does not have this effect. In this example, Pr_1 represents your enduring beliefs about Johnny's travelling habits, however acquired. In contrast, Pr_2 concerns how you revise your beliefs about a specific journey based on this knowledge, in which you learn he travelled to Manchester, or he did not take the train, and so on. So what remains invariant in the revision from Pr_1 to Pr_2 is the target conditional probability, $\Pr(\neg q|\neg p)$ for DA…etc. However, this revision, required for inference, does not mean that people abandon their enduring beliefs about Johnny's travelling habits in Pr_1. Although nothing intrinsic to probability theory enforces this distinction, it is enforced in algorithms for implementing probabilistic inference, for example, Bayes nets.

Bayes nets. A simple Bayes net implementing the JPD Pr_1 in Table 1, consists of two nodes, *p* and *q*, corresponding to Bayesian random variables each with two possible states, 1 (True) and 0 (False), and a directional link from *p* to *q*. Inference over the net consists of variable instantiation, that is, setting *p* or *q* to one of their states, say, $p = 1$, and belief propagation across the link to the *q* node or backwards to the *p* node. The probability that the *q* node is in either of its two states is determined by its conditional probability table (CPT), which includes $\Pr(q = 1|p = 1) = .75$ (and so $\Pr(q = 0|p = 1) = .25$) and $\Pr(q = 1|p = 0) = .167$ (and so $\Pr(q = 0|p = 0) = .833$). Together with the marginal for *p*, $\Pr(p = 1) = .4$, the parameters $\Pr(q = 1|p = 1) = .75$, and $\Pr(q = 1|p = 0) = .167$ implements the JPD Pr_1 in Table 1 in the network. These parameters encode our enduring beliefs about Johnny's travelling habits and remain invariant across different instantiations of its variables to their states.

In this framework, the evidence provided by the categorical premise need not persuade us that, for example, the probability that Johnny travels to Manchester is 1, $\Pr(p) = 1$, and so we should now believe he travels nowhere else. Rather it provides hard evidence to instantiate *p* to 1, and to read off the probability that $q = 1$, in an MP inference. Hard evidence always instantiates a variable to just one of its states. This process is like performing a Ramsey test, supposing the categorical premise by instantiating the relevant state of a random variable, adjusting (i.e., forward and backward belief propagation), and then reading off the probability of the conclusion, which for MP will be the conditional probability $\Pr(q = 1|p = 1)$. This process is the same for the remaining inferences by forward (MP, DA) or backward belief propagation (MT, AC). Like Bayesian conditionalization, it also justifies all our predictions and can be extended to provide an algorithmic level account of inference with contrast sets.

Bayes nets, negative evidence, and contrast sets. We can implement the JPD in Table 2 in a Bayesian network with ternary, rather than binary states, with the CPT in Table 12. This CPT contains two random variables p (travel destinations) and q (modes of transport) with states $\{p_1, p_2, p_3\}$ and $\{q_1, q_2, q_3\}$ respectively. The assertion *Johnny did not travel to Manchester* ($p = \neg p_1$), does not provide hard evidence concerning to which other destination, Paris or Dublin, he did travel. Rather, it provides negative evidence that p can only be instantiated to states p_2 or p_3 but not to p_1 (Bilmes, 2004; Mrad, Delcroix, Piechowiak, Leicester, Mohamed, 2015; Pearl, 1988).

Table 12

Conditional probability table for a Bayes Net with ternary states implementing the JPD in Table 2 showing the conditional probabilities Pr(q_i|p_i) and marginals for p_i.

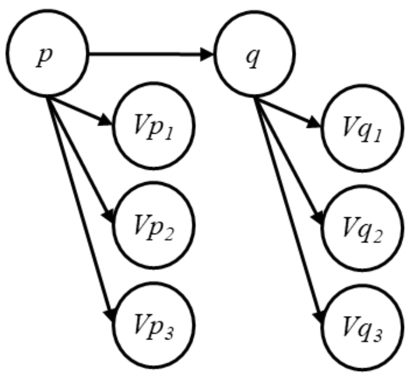
	$p = p_1$ **(.40)**	$p = p_2$ **(.16)**	$p = p_3$**(.44)**
$q = q_1$	0.750	0.625	0
$q = q_2$	0.100	0.250	0.500
$q = q_3$	0.150	0.125	0.500

Note: p_1 = Manchester, p_2 = Paris, p_3 = Dublin, q_1 = train, q_2 = ferry, q_3 = plane.

Following Pearl (1988), we can implement updating on negative evidence using virtual nodes for each state of p and q. These virtual nodes are the children of the ternary nodes p and q in a Bayes net (see Figure 7) with Table 12 as the CPT for the q node (see also, Bilmes, 2004; Mrad et al., 2015). Figure 7 also shows the CPTs for the virtual nodes Vx_y. For the state p_1 of node p $\Pr(Vp_1 = 0|p = p_1) = 0$. Consequently, if $Vp_1 = 0$, then the travel destination (p) cannot be Manchester (p_1), $p \neq p_1$. So the categorical premise *Johnny did not travel to Manchester* provides evidence that $Vp_1 = 0$, and consequently that state p_1 is no longer a possible state of p but that both p_2 and p_3 are possible because $\Pr(Vp_1 = 0|p = p_2) = 1$ and $\Pr(Vp_1 = 0|p = p_3) = 1$. This Bayes net implements exactly the calculations we carried out over the JPD in Table 2 to derive our predictions.[10] Once this Bayes net is learned, inference is easy, and carried out by variable instantiation and belief propagation, without the need for any conscious mental calculation. For example, MP on (1), with the categorical premise *Johnny did not travel to Manchester*, involves instantiating $Vp_1 = 0$, updating the network, and reading off the probability that $Vq_1 = 0$.[11]

Figure 7

Bayes Net implementing the CPT in Table 12 with virtual nodes implementing updating on negative evidence



CPTs for Vx_y ($x = \{p, q\}, y = \{1, 2, 3\}$)

	$x = x_1$		$x = x_2$		$x = x_3$	
	$Vx_y = 1$	$Vx_y = 0$	$Vx_y = 1$	$Vx_y = 0$	$Vx_y = 1$	$Vx_y = 0$
$y = 1$	1	0	0	1	0	1
$y = 2$	0	1	1	0	0	1
$y = 3$	0	1	0	1	1	0

It could be argued that this Bayes net would only work well for small contrast sets. Nonetheless, given that on any particular occasion of using a negation, context and other

[10] It could be argued that this process does not capture the logical inferences that we purport to study. Nonetheless, our experiments, and many others, present participants with versions of the standard logical inference patterns (MP, MT, AC, & DA). Whether or not belief propagation in Bayes nets adequately characterizes these inference patterns from a logical point of view, this process may nonetheless account for how people respond to these inference patterns when presented in experimental tasks and in the real world. Moreover, this may be because people are not particularly interested in what logically follows from some premises, what they want to know is how to update, revise, or otherwise change their beliefs so that they can act appropriately (Harman 1986; Oaksford & Chater, 2020a).

[11] In contrast, calculating $\Pr(\neg q_1|\neg p_1)$ over the JPD in Table 2 involves the following calculation: $(\Pr(p_2, q_2) + \Pr(p_2, q_3) + \Pr(p_3, q_2) + \Pr(p_3, q_3))/(\Pr(p_2, q_1) + \Pr(p_2, q_2) + \Pr(p_2, q_3) + \Pr(p_3, q_1) + \Pr(p_3, q_2) + \Pr(p_3, q_3))$, which we used to derive our predictions.

pragmatic factors will strongly constrain the contrast set, this may be all that is needed (Oaksford & Stenning, 1992). Moreover, as we have argued (see introduction to *Experiment* 2), in inference people only build very limited small-scale generative models related to their immediate deictic or linguistic context (Oakford & Chater, 2020a).[12] These models are constructed on the fly (Chater, 2018) based on linguistic information and prior knowledge, in particular, from immediate past experience, as in decision by sampling models (Stewart, et al., 2006).

The Bayes net in Figure 7 also captures many of our intuitions about contrast sets; in particular, that their internal probabilistic structure will render some contrast set members more likely than others. Take the following examples with the word in bold stressed in speech.

Johnny did not travel to **Manchester** by train (5)

Johnny did not travel to Paris by **train** (5′)

The **cat** was not black (5″)

The cat was not **black** (5‴)

In (5) Johnny travelled somewhere else by train, not Manchester, in (5″) Johnny travelled to Paris by some other mode of transport, not train, in (5″) some other animal was black, not the cat, and in (5‴) the cat was some other colour, not black. Identifying the most likely contrast set member for destination (5) involves instantiating p to $\neg p_1$, on negative evidence, and q to q_1. The model then identifies Paris as the most likely contrast set member, because Pr($p = p_2 | Vp_1 = 0$, $q = q_1$) = 1 and Pr($p = p_3 | Vp_1 = 0$, $q = q_1$) = 0. In (5′), the model identifies ferry as the most likely contrast set member because Pr($q = q_2 | p = p_2$ $Vq_1 = 0$) = .67 but Pr($q = q_3 | p = p_2$, $Vq_1 = 0$) = .33. Directly analogous effects will occur for (5″) and (5‴). These effects suggest that the Bayes net

[12] In this, we agree with mental models theory, although, we disagree on the nature of the small scale models people construct.

in Figure 7 may provide a more general theory of contrary negation and the effects of negative focus in speech.

Causal Bayes nets. We have previously argued that people mentally represent conditionals in causal Bayes nets (Ali, Chater, & Oaksford, 2011; Ali, Schlottman, Shaw, Chater, & Oaksford, 2010; Chater & Oaksford, 2006; Oaksford & Chater, 2010b, 2013, 2016, 2017). However, to capture the implicit negation effect, we have not needed to assume any general probabilistic independencies and so the Bayes net in Figure 7 has been sufficient.[13] However, our account of how people compute contrast sets borrows partly from causal approaches to category structure, in which intrinsic properties of a category cause the various features it possesses (Rehder, 2003a, 2003b, 2017). Moreover, we have suggested that people think about habits like causes, so, for Johnny, travelling to Manchester causes him to travel by train (Oaksford & Chater, 2010, 2020b). We may acquire habits and dispositions from our parents, peers, culture or by intention, but they are rapidly sedimented into the unconscious causes of our actions. All the elements of the ad hoc superordinate category (Barsalou, 1983)—places to which Johnny travels (p)—are causally related to travel destinations considered as features (q). It is a desiderata, therefore, to investigate models integrating CBNs with negative evidence in modelling conditional reasoning.

A minor complication is that if we model contrast sets causally then the direction of causality matters. Some of our materials were diagnostic conditionals, for example, in the vehicles materials the conditional was *if it is not white, then it is not a van.* We think of objects like vans as having features like colour and that it is some intrinsic property of the object that

[13] See *Supplementary Online Materials*: *Section* for an example CBN with parameters corresponding to the JPD in Pr_1 in Table 1.

causes its colour.[14] A CBN representation would require representing the consequent (q) as the cause and the antecedent (p) as the effect. This complication is minor, because we already know from their patterns of discounting and augmentation inferences that people recode diagnostic conditionals in this way (Ali et al., 2011).

A possible argument against the appeal to CBNs, concern recent demonstrations that people violate the independence assumptions of these models (Rehder, 2014; Rottman & Hastie, 2016). However, there are models that can account for these violations (Rehder, 2018). Moreover, the empirically most adequate model may arise from limited sampling from initially preferred states of the underlying generative causal model (Davis & Rehder, 2017; Rehder, 2018). It remains to be seen whether similar violations occur when identifying contrast set members, but the theoretical machinery may be in place to explain them. Processing accounts based on limited sampling from an underlying generative model have also been used to explain away a variety of other biases (Dasgupta, et al., 2017; Hattori, 2016; Sanborn & Chater, 2016; Stewart, et al., 2006)

Alternative Theories

There are three alternative theories of the implicit negations effect, the matching heuristic (Evans, 1998; Thompson, Evans & Campbell, 2013), mental models theory (MMT; Johnson-Laird & Byrne, 2002; Khemlani, Orenes, & Johnson-Laird, 2012), and the cardinality of the

[14] White is the cheapest “vanilla” option that manufacturers provide for vans, and white vans are therefore very common. In the UK, there is even a phenomenon of the “white van driver,” usually fast and discourteous. Consequently, it is a reasonable claim to make that if the vehicle was not white it probably was not a van. Of course, although these are *reasons* for why many vans are white, philosophically reasons are not causes. However, we have argued that people think about most dependencies as if they were causal (Oaksford & Chater, 2010, 2020b).

contrast set hypothesis (Schroyens & Schaeken, 2000; Schroyens, Verschueren, Schaeken, & d'Ydewalle, 2000). MMT implements the double hurdle theory proposed by proponents of the heuristic approach. Consequently, these theories stand, and fall, together. The first hurdle is to see an implicit negation as relevant, that is, as an instance of the negated antecedent or consequent of a conditional.[15] In MMT, negations are represented using explicit contradictory negation tags. The first hurdle is that, unless people can recode the implicitly negated categorical premise using such a tag, they do not realize that a constituent in a mental model has been denied or affirmed. The second hurdle requires a double negation inference, so MT on (1), requires the inference from *it is not the case that he did not travel to Manchester* ($\neg\neg p$) to *he travelled to Manchester* ($\neg\neg p \rightarrow p$). This inference is only required for DA and MT. Both theories locate the problem with implicit negations solely as a difficulty in seeing them as denying or affirming a negated antecedent or consequent. Consequently, they do not predict any of the probabilistic effects we observed.

Binary sets, where there are, say, just two letters {*A, K*} and the contrast set is a singleton, remove the implicit negation effects in comparison to larger sets {*A, K*, *W*} where the contrast set has more than one member (Schroyens, Schaeken, Verschueren, & d'Ydewalle, 2000). The *cardinality of the contrast set* hypothesis (CCS) is that a contrast set with more than one member causes the implicit negation effect. According to this hypothesis with larger contrast sets, participants find it difficult to regard the specific instance, *K*, as representing the superordinate

[15] The matching heuristic describes peoples' apparent inability to deal with mis-matching cases. So, for a conditional, *if A then not 2*, they find it difficult to recognise *K* as denying the antecedent or *7* as affirming the consequent. In Wason's selection task (Evans & Lynch, 1973), this inability leads participants to *match*, that is, they select instances named in the conditional, *A* and *2*, as the cards they need to turn over to verify or falsify it (assuming it describes what is on the faces of double sided cards, of which they can only see one side). Although logically correct for this conditional, they also select *A* and *2* for *if A then 2*.

category, letters that are *not A*. Schroyens et al. (2000) observed implicit negation effects for contrasts sets with two or more members (overall sets sizes of three or more) but not for singleton sets. Although CCS exploits the notion of a contrast set, it does not appeal to their role in computing probabilities. All the contrast sets in our experiments had two members. Consequently, our probabilistic manipulations removed the implicit negation effect even for contrast sets whose cardinalities were greater than one (we refer to this situation as "contrast set(s) > 1"), which is not consistent with the CCS hypothesis. We now briefly consider some recent further evidence supportive of the matching heuristic or mental models.

In the Wason selection task, the matching heuristic response (see, *Footnote* 15) seems meta-cognitively fluent (Thompson, et al., 2013). That is, participants' "answers consistent with a matching heuristic (i.e., selecting cards named in the rule) were made more quickly than other answers, were given higher FOR [feeling of rightness] ratings, and received less subsequent analysis as measured by rethinking time and the probability of changing answers" (Thompson, et al., 2013, p. 431). From a probabilistic perspective, this is not surprising as the probabilistic contrast set account makes the same predictions in this evidence acquisition task (Oaksford & Chater, 2003; 2007; Oaksford, Chater, Grainger & Larkin, 1997). It, therefore, provides a rational analysis of why in data acquisition a matching heuristic is rational. The question of whether this rational analysis is implemented by a heuristic or a probabilistic algorithm depends on whether behaviour can be changed by probabilistic manipulations and the results show that this is possible (e.g., Oaksford et al., 1997). We know of no similar demonstration of fluency for the matching responses in conditional inference. However, we would speculate that if people deploy such a heuristic in the conditional inference task, it is probably learned rather than hard-wired and so can be overridden by subsequent learning, as our experiments demonstrated.

The motivation for an explicit negation tag in MMT derives from the psycholinguistic literature where it is hypothesized that people construct two representations of a negated assertion like “the door is not open” (Kaup, Zwaan, & Lüdtke, 2007; Khemlani et al., 2012, Orenes, Beltran, & Santamaria, 2014). In the first representation, the door is open and in the second, it is closed. This strategy works for binary opposites or antonyms, like open and closed, but what about “the dot is not blue” presented in an array of four coloured dots (Orenes et al. 2014)? Here the second representation would have to include all the other three dots. The negations tag therefore acts as a short hand for the opposites when the overall set size is greater than two. If people represent opposites (contrast sets) for the contrast set > 1 case using a negations tag, then the content of both representations still includes the affirmative statement (e.g., blue dot). Using a visual world array like this, Orenes et al. (2014) used an innovative eye tracking experiment to show that visual attention switches to the alternative when sets are binary (singleton contrast set) but remains on the affirmative item when the contrast set > 1. A finding that is consistent with the use of a negation tag for non-binary opposites.

There are several points to make. First, in these visual world tasks, participants did not have to draw inferences, nothing depended on what the contrast set members might predict. Second, unlike our more real world materials, the contrast sets had no probabilistic structure. So, if the coloured dot was not blue it was equally likely to be one of the other three dots in the display. In our materials, for example, if Johnny did not travel to Manchester, he was far more likely to travel to Dublin than to Paris. Third, our experiments showed that people do not seem to have any trouble representing structured contrast sets with more than one member and drawing appropriate inferences over whatever mental representations of this situation they construct. Fourth, it also seems theoretically incongruous to argue that people automatically recode contrasts sets > 1 with negation tags but also argue that the use of a member of a contrast set > 1 to deny (affirm) a (negated) proposition causes a recoding problem. If people automatically recode these sets with negations tags, then why do they not automatically recode members of one of these sets when encountered in inference? If these contrast sets are *automatically* recoded with a negation tag, then the first hurdle in the mental model implementation of double hurdle theory has been jumped. Moreover, the second hurdle, double negation inferences for MT and DA, is probably a red herring. Our mini meta-analysis showed strong implicit negations effects also for MP and AC (see the introduction to *Experiment 1*), which our experiments replicated.

Although it is unclear how it could integrate with the MMT account of the implicit negation effect, MMTs have been extended to capture probabilistic effects by annotating the possibilities they represent with probabilities (Johnson-Laird, Legrenzi, Girotto, Legrenzi, & Caverni, 1999). To model the current data this would involve representing the nine possible states in the JPDs in Tables 2 and 6 and their associated probabilities. The resulting mental model would be a notational variant of these tables. People would then have to calculate the relevant conditional probabilities by summing over the annotations to the relevant models (cells) and using the ratio formula (see *Footnote 11*). Prima facie, it seems unlikely that people are performing these calculations during inference, rather than compiling a representation as in Figure 7 during learning. Of course, because either theory would predict the same subjective calculated conditional probabilities they would predict the odds of people endorsing an inference equally well. The problem for MMT is that this is not its theory of the implicit negation effect. Moreover, it proposes an implausibly direct implementation of the joint probability distributions in Tables 2 and 6 and of the operations defined over them.

We do not need to deny that our mental representations use negation tags on occasion. As we have pointed out, identifying contrast sets does not exhaust the way people used negations in natural language (Horn, 1989), and some may require people to represent information with a negation tag. We would argue, however, that our normally shallow knowledge of the world (Keil & Rozenblit, 2004; Sloman & Fernbach, 2017), like someone's knowledge of Johnny's travelling habits, means that most contrast sets are not large and are not much like the abstract domains of letters, numbers or coloured dots.

Modelling the Default Prior Pr_0.

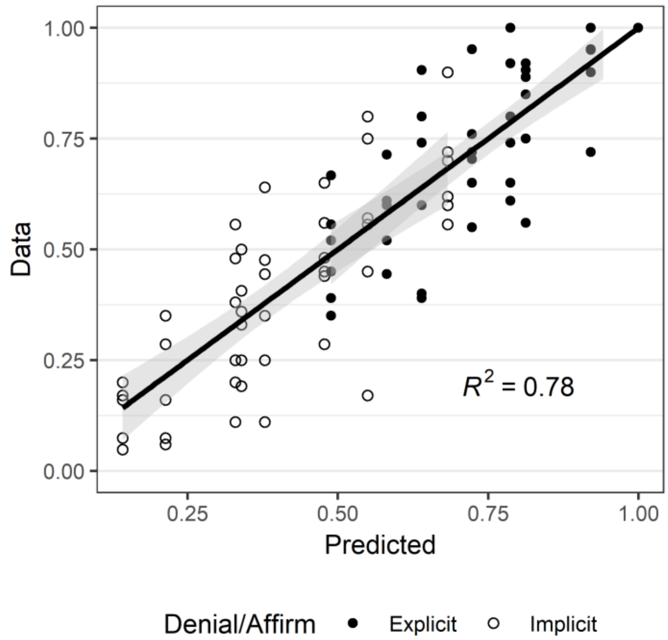
Our focus has been on showing that targeted experimental manipulations of probabilities can produce or remove the implicit negation effect. However, can our account model the original implicit negations effect? The data have been reported in two different ways. Evans and Handley (1999) contrast whole tasks using explicit negations only (the explicit negations paradigm) with whole tasks using implicit negations only (the implicit negations paradigm). Eight of the possible sixteen conditions can reveal implicit negations effects. For example, MP on *if* $\neg p_1$ *then* q_1 can use an explicit, $\neg p_1$, or an implicit, p_2, categorical premise. The implicit paradigm alone also has eight conditions that reveal implicit negations effects (Schroyen et al. 2000). For example, MP on *if* p_1 *then* q_1 must use p_1 to assert the affirmative antecedent, whereas MP on *if* $\neg p_1$ *then* q_1 can use a contrast set member p_2 to assert the negative antecedent. Both cases produce an implicit negations effect. For the same inference (e.g., MP) endorsements of the conclusion (q_1) fall compared to using the explicit negation ($\neg p_1$) on the same rule (*if* $\neg p_1$ *then* q_1) or the affirmative (p_1) on a different rule (*if* p_1 *then* q_1) where the target clause is affirmative. Here we modelled the data from the implicit negations paradigm.

We modelled the six implicit negations paradigm conditions in Evans and Handley (1999: Experiments 1: conditions: no-pictures, pictures, & Experiment 3) and Schroyens et al. (2000: Experiment 1: conditions: set sizes 3, 5, and 9). There were 131 participants and 96 data points. There is one complication. We had to model each of the four rules as if they involved different content. First, this is always the case experimentally because the intention was to see what follows from each rule independently. Second, if the same content is used, as it has been in examples apparently questioning the probabilistic interpretation (Schoyens & Schaeken, 2003), various conceptual absurdities result (Oaksford & Chater, 2003b). Third, the probability conditional does not allow certain pairs of conditionals to be true (or to have high probability) at the same time. The probability conditional respects the law of conditional excluded middle. In standard binary logic *if p then q* and *if p then ¬q* are consistent. They can both be true if the antecedent is false. In contrast, for the probability conditional, for which Pr(*if p then q*) = Pr($q|p$), these conditionals cannot be true together because if Pr($q|p$) = 1, then Pr($\neg q|p$) = 0.[16] So, if these conditionals shared the same content then they cannot both have a high probability. The same argument applies to the pair *if ¬p then q* and *if ¬p then ¬q*. Finally, the four conditionals in the negations paradigm are also related by necessity and sufficiency. So, if they share content, then *if p then q* suggests that *p* is sufficient for *q* and *if ¬p then ¬q* suggests that *p* is necessary for *q*. If *p* is necessary and sufficient for *q* then this should affect endorsements of DA and AC, which would now be valid inferences. In summary, using the same content creates unwanted dependencies between the four conditionals that we can rule out only by using different content as is typically done in these experiments.

[16] However, many advocates of the probability conditional hold that they do not have truth conditions, and, consequently, it would be more accurate to say that these two conditionals cannot both be acceptable.

Figure 8

Modelling the Implicit Negation Effect



We fitted the model using the minimal contrast set structure of two members (overall set size = three) for both antecedent and consequent as in Tables 2 and 6. We modelled each conditional separately thereby assuming different content. The parameters were the nine joint probabilities (a – i), which, because they must sum to one, meant there were eight free parameters, to model 24 data points. Because the data constitute six replications of 16 data points, the best a model can do is predict the mean across replications. With this number of free parameters, this was indeed the outcome of the model fitting (see, Figure 8), the model accounted for 78% of the variance in the data (coefficient to determination R^2 = .78).

Figure 8 also separates out the data points for which a contrast set member (implicit) affirms a negative or denies an affirmative (unfilled dots) and those where the negated constituent (explicit) affirms a negative or denies an affirmative (filled dots). Figure 8 shows that the implicit data and the predicted conditional probability were always lower than the explicit cases. So, the explicit cases (*if p then q, if p then ¬q*) for MP, always had higher probabilities of the conclusion/proportion of endorsements than the implicit cases (*if ¬p then q, if ¬p then ¬q*). We show the best fitting parameter values in the Appendix, Table A1. They will allow us to calculate various quantities to see whether these results conform to recent proposals about conditional inference called "inferentialism."

In summary, our account of the implicit negation effect can account for the original effects observed using all four rules in the negations paradigm. The fundamental insight is that the use of a contrast set member raises the possibility that it does not predict the conclusion as strongly as the explicitly negated categorical premise of a conditional inference. In this sense, the cardinality of the contrast set account is correct in that any contrast set > 1 will raise this possibility (Schroyens, et al., 2000). However, the internal probabilistic structure of the ad hoc categories suggested by the assertion of the conditional causes the effect, not a difficulty in recognizing the contrast set member as an instance of the negated category.

Probabilities

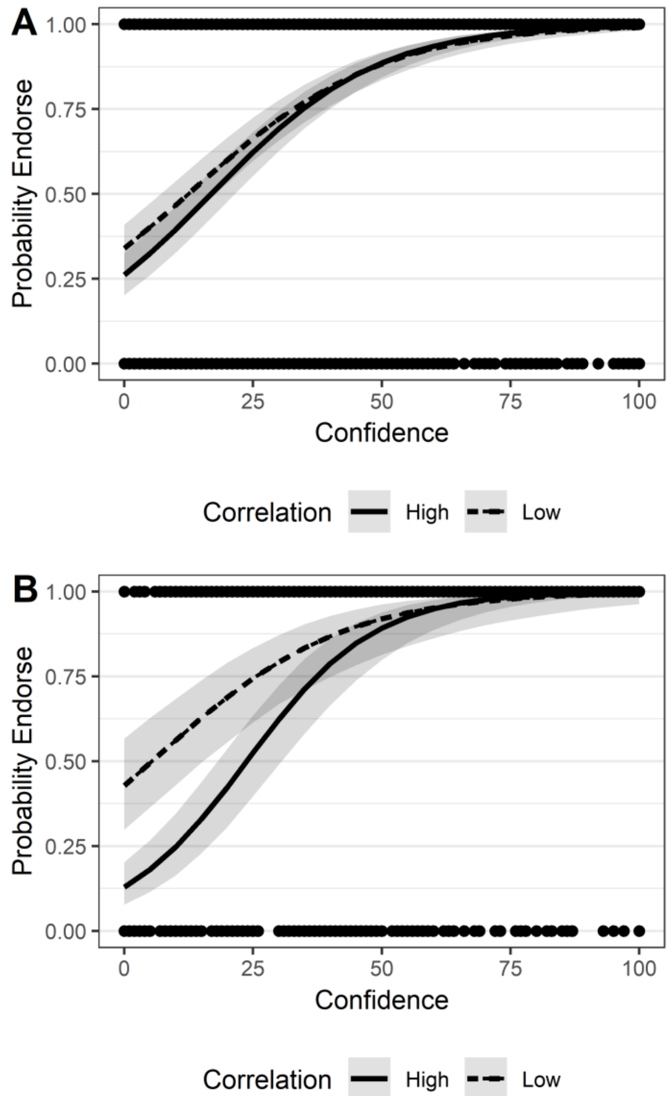
The calculated conditional probabilities predicted the odds of endorsing an inference well. However, even for those participants who understood the probability manipulation (high correlation) very low probabilities still frequently led people to endorse an inference. We could not expect people's subjective probabilities to track the objective probability manipulation exactly. On the Bayesian view of probabilities, they are always relative to what somebody knows or believes, so the general form of a subjective probability statement is Pr(p|B), where B stands for an individual's background beliefs. People know more about the domains of animals and vehicles and their colours than is given in the probability-learning task. Although the subjective estimates did follow the objective probabilities quite well.

One reason why endorsement rates may be high even for low calculated conditional probabilities, is that across all conditions the mean conditional probability was high at around 0.7 (Expt. 1: Objective = .72, Subjective = .68(.19); Expt. 2: Objective = .71, Subjective = .75(.32); Expt. 3: Objective = .71, Subjective = .75(.32)). Consequently, on average, participants should endorse an inference, although this will depend on their personal criterion or cut-off. Moreover, they should endorse five out the six inferences they experienced in each manipulation, which again may bias participants towards endorsement. Given this potential bias toward endorsement, it is impressive that our results nonetheless showed a strong effect of calculated conditional probability on the odds of endorsing an inference.

Another reason why the calculated conditional probabilities may not be better predictors of inference endorsement is the indirect method of computation and the reliance on the ratio formula to compute the conditional probabilities ($\Pr(q|p) = \Pr(p, q)/\Pr(p)$). The probability verification task is similar to versions of the probabilistic truth table task (Over et al, 2007). This task has been criticized as perhaps not revealing people's probabilistic interpretations of the conditional (Jubin & Barrouillet, 2019). The precise reasons do not matter, but an immediate response is that (a) these tasks (especially our task which involves filling in 9 cells of the JPD) creates a lot of room for error, and (b) the subjective Bayesian approach rejects the frequentist method and the ratio formula for calculating conditional probabilities. On the Bayesian interpretation, conditional probabilities are basic and suppositional, that is, they based on the Ramsey test (see, *Probabilities and Contrast Sets*).

Figure 9

Predicting Endorsement Rates from Confidence for High and Low Correlation Groups



*Notes: A: Experiment 2, B: Experiment 3. For both experiments the model fitted was Endorse ~ Conf*Corr + (1|Participant) + (Conf *Corr|PaGr).*

People's probability judgements are more coherent when queried while drawing inferences (Evans, Thompson, & Over, 2015). We have already shown that in our experiments, calculated conditional probability directly predicts confidence in endorsing an inference. Therefore, people's confidence judgements, which we obtained when people are actually drawing inferences, may provide a more direct measure of the relevant conditional probabilities. As we have argued, during inference people effectively perform a Ramsey test, supposing the categorical premise to be true (see, *Bayes nets*). If their degree of belief in the conclusion goes above criterion, then they endorse the inference and report this degree of belief as how confident they are. If this is the right interpretation, then the suppositional account would predict that using confidence as a predictor should lead to a much steeper response curve showing sensitivity at both the high and the low ends of the scale. Moreover, if the probability-learning task has influenced people's subjective conditional probabilities as measured by the confidence judgements, then we would expect to see a moderating effect of high or low correlation (*Corr*).

Figure 9 shows how the odds of endorsing an inference varied with confidence for the high and low correlation groups in Experiments 2 and 3. As predicted, the response curves are much steeper than for calculated conditional probability, and correlation in the probability verification task moderated the effect, especially in Experiment 3. Table 13 shows that in both Experiments 2 and 3, using confidence (M1) as a predictor yielded a much better fit to the data than calculated conditional probability (M2). However, even in the high correlation group in Experiment 3, people still seem biased to endorse an inference as revealed by the left-shift in the response curve (see, Figure 9). One would expect the odds of endorsing an inference to be one (probability = 0.5) when conditional probability was 0.5. As we observed, this may be because, on average, inferences in this task should be endorsed. De-biasing may be possible by balancing inferences so that equal numbers should be endorsed or rejected. The moderating effect of correlation demonstrates that the effects of the learning-phase endured to affect people's subjective probability judgements, as measured by confidence, in the inference tasks.

Table 13

Model Comparison for Predicting Inference Endorsement Rates from Confidence vs. Calculated Conditional Probability

	LOOIC	*SE*	*k*	*ΔLOOIC*	*Δelpd*	*Δse*	*Weight*
Experiment 2							
M1	1852.0	75.2	6.6	0	0	0	.72
M2	2170.3	75.7	5.8	318.3	-159.2	33.4	.28
Experiment 3							
M1	788.4	50.9	6.1	0	0	0	.73
M2	930.2	50.7	5.7	141.8	-70.9	23.2	.27

Notes. M1: Confidence, M2: Calculated Conditional Probability. Estimated number of parameters (k), the difference (ΔLOOIC), the difference in expected log posterior predictive density (Δelpd) and its standard error (Δse), and the Bayesian stacking weights (LOOIC-weight).

Inferentialism

A recent development in the psychology of reasoning is the realization that people tend to endorse conditionals only when they believe there is some kind of inferential link between the antecedent and the consequent. So for example, they do not regard conditionals like, *if the moon is made of cheese, Corbyn will be elected Prime Minister* as candidates for truth. Although, given that the moon is not made of cheese, we would logically have to endorse this conditional as true. This is one of the so-called “paradoxes of material implication.” There are two versions of inferentialism. On the semantic version, indicative conditionals express inferential or reason relations between the antecedent and consequent which are part of the truth conditions of the conditional (Douven, Elqayam, Singmann, & van Wijnbergen-Huitink, 2018; Douven & Mirabile, 2018; Mirabile & Douven, in press). On the probabilistic version reason relations are probabilistic and part of the acceptability conditions of indicative conditionals (Krzyżanowska, Collins, & Hahn, 2017; Skovgaard-Olsen, Collins, Krzyżanowska, Hahn, & Klauer, 2019; Skovgaard-Olsen, Kellen, Hahn, & Klauer, 2019; Skovgaard-Olsen, Kellen, Krahl, & Klauer, 2017; Skovgaard-Olsen, Singmann, & Klauer, 2016, 2017). Antecedent and consequent are positively probabilistically relevant when $\Pr(q|p) > \Pr(q|\neg p)$, that is, when Delta-P (ΔP, Ward & Jenkins, 1965) is positive. ΔP was found to moderate whether the Equation ($\Pr(\textit{if } p \textit{ then } q) = \Pr(q|p)$) holds. Only when $\Delta P > 0$, that is, *p* and *q* are positively inferentially relevant, does the Equation adequately predict whether a conditional is acceptable.

The data from the probability verification task and the best fitting parameter values from the model fits (see, *Modelling the default prior* Pr_0) allow us to check whether the materials in these tasks show positive relevance. For Experiment 2, the objective probabilities for the *if ¬p, then ¬q* rule respected positive relevance. For the MP-manipulation, ΔP ($\Pr(\neg q|\neg p) - \Pr(\neg q|p)$) = .91 , and for the AC-Manipulation, $\Delta P = .80$. Aggregating across manipulations, for the subjective probabilities, mean $\Delta P = .64$ ($SD = .36$). Only 54 out of 668 calculated ΔPs (7.8%) were zero or negative and 52 of these came from the low correlation group. For Experiment 3, the objective probabilities for the *if p then q* rule respected positive relevance. For both the DA- and the MT-manipulations, ΔP ($\Pr(q|p) - \Pr(q|\neg p)$) = .80. Aggregating across manipulations, for the subjective probabilities, mean $\Delta P = .51$ ($SD = .46$). 59 out of the 336 calculated ΔPs (17.6%) were zero or negative and all came from the low correlation group. We also checked the best fitting parameter values for the four rules in the implicit negations paradigm task and they also all showed positive relevance (*if p then q*: $\Delta P = .43$; *if p then ¬q*: $\Delta P = .11$; *if ¬p then q*: $\Delta P = .19$;

if ¬p then ¬q: $\Delta P = .09$). It would appear that for abstract conditionals (implicit negations paradigm) and those used in these experiments, people assume positive relevance between antecedent and consequent.

Our results are relevant to an ongoing debate over the truth or acceptability conditions of conditionals. On the suppositional view of the conditional, judging whether a conditional is true or acceptable should depend on the conditional probability. According to semantic inferentialism (Douven, et al., 2018), in addition people must believe that there is an inferential link between antecedent and consequent. The existence of this inferential link explains why the antecedent *explains* the consequent for *if you turn the key the car starts*, but the antecedent of *if the moon is made of cheese, Corbyn will be elected Prime Minister* does not explain the consequent. Another example is the contrast between *if the sun rises, then the cock crows* and *if the cock crows then the sun rises*. Only in the former does the antecedent explain the consequent.[17] This hypothesis has been tested by asking people how well the antecedent of an abductive or diagnostic conditional (e.g, *if the cock crows then the sun rises*) is explained by its consequent (Mirabile & Douven, in press: Experiment 3), thereby providing a measure of explanation quality. Participants also judged how strongly they believed the truth of the conclusion of an MP inference using the same abductive conditionals. Finally, they completed a probabilistic truth table task to obtain a measure of conditional probability. Explanation quality was a better predictor of how strongly someone believed that the conclusion of the MP inference was true than conditional probability. Explanation quality and conditional probability were also correlated, indeed they were more correlated than either was individually with truth.

[17] Although, the inverse could be regarded as an abductive inferential link (Krzyżanowska, Wenmackers, & Douven, 2013).

In looking at the relation between confidence and inference endorsement in the last section, we interpreted the fact that calculated conditional probability and confidence were highly correlated as indicating that confidence provided a more direct measure of conditional probability. That was why confidence was a better predictor of inference endorsement. The same argument applies to Mirabile and Douven's (in press; see also, Douven & Mirabile, 2018) measure of explanatory goodness, which they also assessed directly for each conditional. Consequently, explanatory goodness and confidence may just be better more direct measures of conditional probability than the probabilistic truth table task because they more closely follow the Ramsey test. So, contradicting Mirabile and Douven (in press), a construct of explanatory goodness distinct from conditional probability may not be required to explain the data.

However, although this is a plausible line of argument, we would suggest that when you believe a conditional you believe it describes some underlying, usually causal, dependency in the world (Oaksford & Chater, 2010, 2017, 2020a, 2020b), which is why we suggested modelling these data using causal Bayes nets may be a fruitful line of research. That ΔP was positive for the main conditionals in our experiments showed that people believed the antecedent was positively causally relevant to the consequent because ΔP is the numerator of causal power (Cheng, 1997), which provides the weights on the links in a CBN (see *Supplementary Online Material*). Consequently, like semantic inferentialism, we would argue that the reason why confidence and explanation quality are better predictors of the odds of endorsing an inference is that people directly consider the causal or inferential link, which they do not need to do in the probabilistic truth table task. Indeed, if they learn a Bayes net during the learning phase, which requires them to consider the inferential link and its direction, then it would be difficult to reconstruct the individual cell values of the JPD in the probability verification task. It would require recording the prior over p, instantiating p to each of p_{1-3} and reading off the nine conditional probabilities $\Pr(q = q_{1-3} | p = p_{1-3})$ and multiplying them by the priors $\Pr(p = p_{1-3})$. That people seem capable of doing something like this with some degree of accuracy in the probability verification task is quite impressive. However, we learn about the world in order to predict and explain it and we argue that this requires setting up mental representations that facilitate inference, like the Bayes net in Figure 7.

Learning

Our probability manipulations used brief experiential learning phases, shown in research in judgement and decision making to improve performance (Hogarth & Soyer, 2011; Wulf, et al., 2018). It is worth emphasizing that these learning experiences were short, only 30 trials in Experiments 2 and 3, and no attempt was made to get participants to learn the distributions to any criterion of accuracy. Nonetheless, these learning experiences profoundly influenced participants' behavior when presented with verbal conditional inference problems. All other theories attribute the implicit negations effect to errors in constructing a mental representation of the logical form of the premises. In contrast, we have argued that conditionals describe the dependencies in the world that allow us to predict and explain it (e.g., Oaksford & Chater, 2010, 2020b). It should not be surprising that people are adept at rapidly acquiring the information they need from their immediate environment to build small scale models that allow them to do this and so to act in that environment.

The importance of sampling from the environment is also emphasized in decision by sampling models (Sanborn & Chater, 2016; Stewart, et al. 2006). Samples may be derived from memory, but in novel contexts, where previous experience is little guide, people must sample from the environment. Moreover, the structure of samples or choice options can strongly influence decision making (Stewart, Chater, Stott, & Reimers, 2003). Models like Bayes nets, include information about structure (directed links and independence relations) and strength (causal strength or the relevant CPT). The probabilities that are used to compute strength can come from memory or, in novel contexts, must be sampled from the immediate environment. In Bayes nets there also are algorithms for learning not just the relevant probabilities but also the network structure of these models (Korb & Nicolson, 2010). That is, learning is integral to these models, in a way that it is not in other non-probabilistic theories of verbal reasoning. Moreover, as we have seen, how well participants learned the distribution strongly moderated the effect of calculated conditional probability and confidence on the odds of endorsing as inference.

It could be argued that the reliance of our account, and its implementation in Bays nets, on learning is a limitation as it only applies when probabilities are learned. However, we have shown that the contrast set model also fits the base-line implicit negation effect (see, *Modelling the Default Prior* Pr_0). So the same model applies whether the probabilities are provided by memory or learned from the immediate environment. Although, of course, the default prior was also, presumably, learned, at least in part, from experience. Other probabilistic manipulations may be less effective in producing the discriminatory effects we observed in these experiments. So, Experiment 1 only showed minimal changes to the default prior when participants were given descriptions of the distribution in Table 2 as single event probabilities (e.g., 0.8 or 80%) in the pre-learning inference task. Single event probabilities, it would appear, do not update people's default-priors as effectively as experience, as many have argued (e.g., Gigerenzer & Hoffrage, 1995). However, it remains to be seen if frequency formats (80 out of a 100) (Gigerenzer & Hoffrage, 1995), lead to a more effective update as observed in some previous research (Oaksford, et al., 1997, 1999). Sample summaries (Hawkins et al., 2015) are closely related to frequency formats. It would be interesting to see whether sample summaries of the parameters of the CPT in Table 12 could produce similar effects. These distributions are the most relevant to inference but they relate directly only to the forward inferences (MP and DA). An interesting prediction of the Bayes net implementation is that when presented with only these samples, the backwards inferences (AC and MT) should still track the inverse conditional probabilities.

Rationality

Is people's behavior on these tasks rational? Answering this question depends on what you think people should do when confronted with these inference tasks. Clearly, people are not rational with respect to standard conditional logic. Regardless of the whether the negation in the categorical premise is explicit or implicit, all that is logically relevant is whether it affirms or denies the antecedent or consequent. If it affirms the antecedent (MP) or denies the consequent (MT), the inference should be endorsed otherwise it should not be endorsed. Clearly, people are not rational with respect to this standard as they happily reject inferences when a clause is denied (affirmed) implicitly that they happily accept when it is denied (affirmed) explicitly.

People can *deduce* probabilistic conclusions from uncertain premises (Cruz, Baratgin, Oaksford, & Over, 2015; Evans, Thompson, & Over, 2015; Pfeifer & Kleiter, 2009; Politzer & Baratgin, 2016; Singmann, Klauer, & Over, 2014). In coherence-based probability logics (Coletti & Scozzafava, 2002), we can deduce a probability interval from the probabilities of the major and minor premise. So, for example, suppose that in Experiments 1 and 2 $\Pr(\neg q|\neg p) = 0.8$ and $\Pr(\neg p) = .8$, then the probability of the conclusion of MP must lie in the interval $.64 \geq \Pr(\neg q)$

≤ .84. These intervals respect probabilistic coherence assuming only the information given in the premises. From this probabilistic logic point of view, again the only significance an implicit negation has is being an instance of the relevant negated category. In this paper, we have interpreted the evidence given by the categorical premise as either hard (affirmative) or virtual (negations) evidence concerning the states of the random variables in a Bayes net, which includes full knowledge of the JPD. Probability logic does not typically assume full knowledge of the JPD but allows for uncertainty in the categorical premise. Take for example AC, and assume that the probability of each categorical premise is the relevant marginal probability in Table 2. According to probabilistic coherence, for the explicit negation (AC-Not) the probability of the conclusion of this inference on (1) should be in the interval [0, .278] and for implicit negation (AC-Con) it should be [0, .937]. However, the mean computed conditional probabilities and probabilities of endorsement (in brackets) of each inference was AC-Not: .79 (.97) and AC-Con: .77 (.94). For AC-Not both probabilities fell well outside of the coherence interval. Consequently, people's behavior in these experiments is not rational with respect to the standards of coherence-based probability logic.[18]

From our perspective, reasoning is about rational change of belief (Eva & Hartmann, 2018; Harman, 1986; Oaksford & Chater, 2007, 2020a). Here we have modelled inference as belief propagation or update in Bayes nets, which respect the laws of probability theory. The

[18] It remains possible that probability logic can predict these results by including the information in the learning trials as additional premises. However, to explain the implicit negation effects would seem to require an account of contrary negation, unavailable logically, but readily implemented using virtual nodes in the Bayes net in Figure 7 (Pearl, 1988).

extent to which the relevant conditional probabilities predict inference endorsements show the extent to which we can view peoples' reasoning as rational. In our experimental tasks, the learning samples were taken from the same population of experiences as the informant (e.g., the vet) asserting the conditional, so the premises should not lead to any changes in the probabilities that define people's enduring beliefs in the CPT of their Bayes net representation.. However, there are situations where learning the premises suggests revisions to our degree of belief in a conditional premise (Oaksford & Chater, 2007, 2013). Such situations seem to require revising our beliefs not just updating them supposing the categorical premise is true. Although beyond the scope of our current discussion, guaranteeing the rationality of inference in these dynamic contexts remains a more challenging problem (Douven & Romeijn, 2011; Eva & Hartmann, 2018, Hartmann & Rafiee Rad, 2012; Oaksford & Chater, 2013).

Common Mechanisms

In explaining our results, we have not appealed to any mechanisms that are unique to deductive reasoning. Rather we have argued that mechanisms like Bayes nets may provide an account of the representations and processes underlying the implicit negation effect by providing an implementation of how people learn, represent and access contrast sets. We have previously argued that CBNs may provide an account of conditional inference, not just with causal conditionals (Ali et al., 2011), but with conditionals generally (Oaksford & Chater, 2010a,b). We have also argued that they may provide an implementation of inferentialism (Oaksford & Chater, 2020b). More generally, we have argued that common mechanisms may underlie, inductive, deductive and causal reasoning and these are likely to be similar in kind to those that underlie judgement and decision-making (Oaksford & Chater, 2020a). Proposals for closer relations between deductive inference and other areas of higher cognition are not new: with judgement and decision-making (Manktelow & Over, 1991) and with causal reasoning (Oaksford & Chater, 1994).

However, there is a contrast with the mental models approach, which also provides explanations of inductive, deductive, and causal reasoning (Johnson-Laird, Goodwin, & Khemlani, 2018; Johnson-Laird, & Khemlani, 2017). Mental models treats discrete representations of possibilities as basic. These possibilities are closely related to the truth table cases allowed by the binary logical connectives, but they can be modulated by prior knowledge or labelled to capture other forms of inference. Following many other areas of perception and cognition, we regard the mind/brain's task to be the extraction of useful regularities from the flux of experience in order to predict and ultimately explain the world. The fundamental mode of representation is probabilistic and continuous, and it is only by sampling the brain's underlying stochastic models that we come to represent discrete possibilities. Usually these are just the deliverances to consciousness of the results of the processes that actually drive our behavior. If we do anything more with them it seems as likely to lead to error as to successful reasoning. So, while there is agreement on common mechanism, the new paradigm in reasoning generalizes in the opposite direction to mental models, from other areas of cognition to deduction and not from accounts of deductive reasoning elsewhere.

Conclusion

Psychologists are beginning to uncover the rational basis for many of the biases discovered over the last 50 years in deductive and causal reasoning, judgement and decision-making. In this paper, we have argued that using a manipulation, experiential learning, shown to be effective in judgement and decision-making may elucidate the rational underpinning of the implicit negation effect in conditional inference. In three experiments, we created and removed the effect by using probabilistically structured contrast sets acquired during a brief learning phase. No other theory of the implicit negations effect makes these predictions. We could model our findings well using Bayes nets similar to causal approaches to category structure, which also captured further intuitions about how contrast sets can identify the most likely opposites. We also showed that our results and our Bayes net approach aligns closely to a recent development in the psychology of reasoning called inferentialism. A key feature is that we have not appealed to any cognitive mechanism or module whose specific task is logical reasoning. This approach is consistent with the conclusion of our recent review of new paradigm probabilistic theories, which treats argumentation, deduction and induction alike within a probabilistic framework similar in kind to processes involved in other areas of cognition (Oaksford & Chater, 2020a).

Context

We have been explaining biases in human deductive reasoning using Bayesian rational analysis for 25 years (Oaksford & Chater, 1994, 2020a). This pattern of explanation had seemed paradoxical because Bayesian reasoning in judgement and decision-making had always seemed similarly biased. Recently, however, it has been shown that people's judgement and decision-making can be surprisingly rational when probabilities and utilities are learned by experience. We used experiential learning phases to allow participants to acquire information about probability distributions that should create and remove the implicit negation effect in conditional reasoning. This is the first time that discrete experiential learning has been used to manipulate probabilities in deductive reasoning tasks. We had already shown that our Bayesian approach could rationally explain polarity biases in conditional inference using the concept of a contrast set. Our current experiments show that this account generalises to the implicit negations effect. We could also model the effects well using Bayes nets. We show how these data also apply directly to recent inferentialist accounts of conditional inference. Our results suggest that similar cognitive mechanisms may underlie causal, inductive and deductive reasoning as proposed in our recent review of the new paradigm in the psychology of reasoning (Oaksford & Chater, 2020a).

References

Ali, N., Chater, N., & Oaksford, M. (2011). The mental representation of causal conditional inference: Causal models or mental models. *Cognition*, *119*, 403-418.

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

Anderson, J. R. (1995). *Cognitive psychology and it implications.* New York: W. H. Freeman.

Anderson, J. R., & Sheu, C.-F. (1995). Causal inferences as perceptual judgements. *Memory & Cognition*, *23*, 510–524. https://doi-org.ezproxy.lib.bbk.ac.uk/10.3758/BF03197251

Anderson, S. K., Olesen, K. G., & Jensen, F. V. (1990). HUGIN—a shell for building Bayesian belief universes for expert systems. In G. Shafer & J. Pearl (Eds.), *Readings in uncertain reasoning* (pp. 332-337). San Francisco, CA: Morgan Kaufmann Publishers Inc.

Ardia, D., Mullen, K. M., Peterson, B. G., & Ulrich, J. (2016). DEoptim: Differential Evolution in 'R'. version 2.2-4.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 340–412. http://dx.doi.org/10.1016/j.jml.2007.12.005

Bakdash, J. Z., & Marusich, L. R. (2017). Repeated measures correlation. *Frontiers in Psychology*, *8*, 456. doi: 10.3389/fpsyg.2017.00456

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition*, *11*, 211–227. https://doi-org.ezproxy.lib.bbk.ac.uk/10.3758/BF03196968

Barsalou, L. W., Huttenlocher, J., & Lamberts, K. (1998). Basing categorization on individuals and events. *Cognitive Psychology*, *36*, 203–272. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1006/cogp.1998.0687

Bilmes, J. (2004). *On virtual evidence and soft evidence in Bayesian networks* (UWEE Technical Report No. UWEETR-2004-0016). Seattle, Washington: Department of Electrical Engineering.

Byrne, R. M. J., & Johnson-Laird, P. N. (2019). If and or: Real and counterfactual possibilities in their truth and probability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1037/xlm0000756.supp (Supplemental)

Chater, N. (2018). *The mind is flat*. London: Allen Lane.

Chater, N., & Oaksford, M. (2006). Mental mechanisms. In K. Fiedler, & P. Juslin (Eds.), *Information sampling and adaptive cognition* (pp. 210-236). Cambridge: Cambridge University Press.

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

Chow S, Shao J, Wang H. (2008). *Sample Size Calculations in Clinical Research*. 2nd Ed. Chapman & Hall/CRC Biostatistics Series

Coletti, G., & Scozzafava, R. (2002). *Probabilistic logic in a coherent setting*. Dordrecht, NL: Kluwer.

Costello, F., & Watts, P. (2014). Surprisingly rational: Probability theory plus noise explains biases in judgement. *Psychological Review*, *121*, 463–480. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1037/a0037010

Costello, F., Watts, P., & Fisher, C. (2018). Surprising rationality in probability judgement: Assessing two competing models. *Cognition*, *170*, 280–297. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1016/j.cognition.2017.08.012

Cruz, N., Baratgin, J., Oaksford, M., & Over, D. E. (2015). Reasoning with if and ands and ors. *Frontiers in Psychology*, *6*, 192. doi: 10.3389/fpsyg.2015.00192

Dasgupta, I., Schulz, E., & Gershman, S. J. (2017). Where do hypotheses come from? *Cognitive Psychology*, *96*, 1–25. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1016/j.cogpsych.2017.05.001

Davis, Z., & Rehder, B. (2017). The causal sampler: A sampling approach to causal representation, reasoning and learning. *Poster presented at the 39th annual conference of the Cognitive Science Society*. London, UK.

De Martino, B., Fleming, S. M., Garrett, N., & Dolan, R. J. (2013). Confidence in value-based choice. *Nature Neuroscience*, *16*, 105–110. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1038/nn.3279

Douven, I., Elqayam, S., Singmann, H., & van Wijnbergen-Huitink, J. (2018). Conditionals and inferential connections: A hypothetical inferential theory. *Cognitive Psychology*, *101*, 50–81. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1016/j.cogpsych.2017.09.002

Douven, I., & Mirabile, P. (2018). Best, second-best, and good-enough explanations: How they matter to reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*, 1792–1813. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1037/xlm0000545

Douven, I., & Romeijn, J.-W. (2011). A new resolution of the Judy Benjamin problem. *Mind*, *120*, 637–670.

Edgington, D. (1995). On conditionals. *Mind, 104,* 235–329. http://dx.doi.org/10.1093/mind/104.414.235

Eva, B., & Hartmann, S. (2018). Bayesian argumentation and the value of logical validity. *Psychological Review*, *125*, 806–821. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1037/rev0000114.supp (Supplemental)

Evans, J. St.B. T. (1998). Matching bias in conditional reasoning: Do we understand it after 25 years? *Thinking and Reasoning, 4,* 45–82.

Evans, J.St.B.T. (2002). Matching bias and set sizes: A discussion of Yama (2001). *Thinking & Reasoning*, *8*, 153-163.

Evans, J.St.B.T. (1972). Interpretation and “matching bias” in a reasoning task. *Quarterly Journal of Experimental Psychology*, *24*, 193–199.

Evans, J. St.B. T., Clibbens, J., & Rood, B. (1995). Bias in conditional inference: Implications for mental models and mental logic. *Quarterly Journal of Experimental Psychology*, 48A, 644–670. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1080/14640749508401409

Evans, J.St.B.T., Clibbens, J., & Rood, B. (1996). The role of implicit and explicit negations in conditional reasoning bias. *Journal of Memory & Language*, *35*, 392–409.

Evans, J.St.B.T., & Handley, S.J. (1999). The role of negation in conditional inference. *Quarterly Journal of Experimental Psychology*, *52*, 739–770.

Evans, J. St. B. T., Handley, S. J., & Over, D. E. (2003). Conditionals and conditional probability. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 321–335. http://dx.doi.org/10.1037/0278-7393.29.2.321

Evans, J.St.B.T., & Lynch, J.S. (1973). Matching bias in the selection task. *British Journal of Psychology*, *64*, 391–397.

Evans, J. St.B. T., Thompson, V., & Over, D. E. (2015). Uncertain deduction and conditional reasoning. *Frontiers in Psychology*, *6*, 398.

Fernbach, P. M., & Erb, C. D. (2013). A quantitative causal model theory of conditional reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 1327-1343. doi:10.1037/a0031851

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*, 635–648. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1037/a0022329

Gelman, A. (2018). You need 16 times the sample size to estimate an interaction than to estimate a main effect. Retrieved from: https://statmodeling.stat.columbia.edu/2018/03/15/need-16-times-sample-size-estimate-interaction-estimate-main-effect/

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. New York: CRC Press.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102*, 684-704.

Griffths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology, 51*, 334–384.

Goodrich, B., Gabry, J., Ali, I., Brilleman, S. (2018). rstanarm: Bayesian applied regression modeling via Stan. R package version 2.17.4, http://mc-stan.org/.

Hahn, U., & Oaksford, M. (2007). The rationality of informal argumentation: A Bayesian approach to reasoning fallacies. *Psychological Review*, *114*, 704-732.

Harman, G. (1986). *Change in view: Principles of reasoning*. Cambridge, MA: MIT

Press/Bradford Books.

Hartmann, S., & Rafiee Rad, S. (2012). Updating on conditionals = Kullback-Leibler distance + causal structure. Presented at the *Biennial Meeting of the Philosophy of Science Association*, San Diego, CA, November 15–17.

Hattori, M. (2016). Probabilistic representation in syllogistic reasoning: A theory to integrate mental models and heuristics. *Cognition*, *157*, 296–320. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1016/j.cognition.2016.09.009

Hattori, M., & Oaksford, M. (2007). Adaptive non-interventional heuristics for co-variation detection in causal induction: Model comparison and rational analysis. *Cognitive Science*, *31*, 765-814.

Hawkins, G. E., Hayes, B. K., Donkin, C., Pasqualino, M., & Newell, B. R. (2015). A Bayesian latent-mixture model analysis shows that informative samples reduce base-rate neglect. *Decision*, *2*, 306–318. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1037/dec0000024.supp (Supplemental)

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science*, *15*, 534 –539. http://dx.doi.org/10.1111/j.0956-7976.2004.00715.x

Hinterecker, T., Knauff, M., & Johnson-Laird, P. N. (2016). Modality, probability, and mental models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42,* 1606–1620. http://dx.doi.org/10.1037/xlm0000255

Hogarth, R. M., & Soyer, E. (2011). Sequentially simulated outcomes: Kind experience versus non-transparent description. *Journal of Experimental Psychology: General*, *140*, 434. http://dx.doi.org/10.1037/a0023265

Horn, L. R. (1989). *A natural history of negation*. Chicago: University of Chicago Press.

Jarvstad, A., Hahn, U., Rushton, S. K., & Warren, P. A. (2013). Perceptuo-motor, cognitive, and description-based decision-making seem equally good. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, *110*, 16271–16276. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1073/pnas.1300239110

Johnson-Laird, P. N. (1983). *Mental models*. Cambridge, UK; Cambridge University Press

Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review*, *109*, 646–678. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1037/0033-295X.109.4.646

Johnson-Laird, P. N., Goodwin, G. P., & Khemlani, S. S. (2018). Mental models and reasoning. In L. J. Ball & V. A. Thompson (Eds.), *The Routledge international handbook of thinking and reasoning*. (pp. 346–365). New York, NY: Routledge/Taylor & Francis Group.

Johnson-Laird, P. N., & Khemlani, S. S. (2017). Mental models and causation. In M. R. Waldmann (Ed.), *The Oxford handbook of causal reasoning*. (pp. 169–187). New York, NY: Oxford University Press.

Johnson-Laird, P. N., Legrenzi, P., Girotto, V., Legrenzi, M. S., & Caverni, J. P. (1999). Naïve probability: A mental model theory of extensional reasoning. *Psychological Review*, *106*, 62–88.

Joseph, L., du Berger, R., & Belisle P. (1997). Bayesian and mixed Bayesian/likelihood criteria for sample size determination propdiff.acc 3. *Statistics in Medicine*, *16*, 769-781.

Jubin, J., & Barrouillet, P. (2019). Effects of context on the rate of conjunctive responses in the probabilistic truth table task. *Thinking & Reasoning*, *25*, 133–150. https://doi.org/10.1080/13546783.2018.1477689

Kay, M. (2019). tidybayes: Tidy data and geoms for Bayesian models. doi: 10.5281/zenodo.1308151 (URL: http://doi.org/10.5281/zenodo.1308151), R package version 1.0.4, <URL: http://mjskay.github.io/tidybayes/>.

Khemlani, S., Hinterecker, T., & Johnson-Laird, P. N. (2017). The provenance of modal inference. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. J. Davelaar (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (pp. 663–668). Austin, TX: Cognitive Science Society.

Khemlani, S., Orenes, I., & Johnson-Laird, P. N. (2012). Negation: A theory of its meaning, representation, and use. *Journal of Cognitive Psychology*, *24*, 541–559. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1080/20445911.2012.660913

Kleiter, G. D., (1994). Natural sampling: Rationality without base rates. In G. H. Fischer & D. Laming (Eds.), *Contributions to mathematical psychology, psychometrics, and methodology* (pp. 375-388). New York: Springer.

Korb, K. B., & Nicholson, A. E. (2010). *Bayesian artificial intelligence.* New York: CRC Press.

Kruschke, J. (2011). *Doing Bayesian data analysis*. New York: Academic Press.

Krzyżanowska, K., Wenmackers, S., & Douven, I. (2013). Inferential conditionals and evidentiality. *Journal of Logic, Language and Information*, *22*, 315–334. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1007/s10849-013-9178-4

Lenth, R. (2019). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.3.3. https://CRAN.R-project.org/package=emmeans

Lüdecke, D. (2018). sjPlot: Data visualization for statistics in social science. doi: 10.5281/zenodo.1308157 (URL: http://doi.org/10.5281/zenodo.1308157), R package version 2.6.2, <URL: https://CRAN.R-project.org/package=sjPlot>.

Marcus, S. L., & Rips, L. J. (1979). Conditional reasoning. *Journal of Verbal Learning & Verbal Behavior*, *18*, 199–223. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1016/S0022-5371(79)90127-0

McElreath, R. (2016). *Statistical rethinking.* New York: CRC Press.

McKenzie, C. R. M. (2004). Framing effects in inference tasks and why they are normatively defensible. *Memory and Cognition*, *32,* 874-885.

McKenzie, C. R. M., Ferreira, V. S., Mikkelsen, L. A., McDermott, K. J., & Skrable, R. P. (2001). Do conditional statements target rare events? *Organizational Behavior & Human Decision Processes*, *85*, 291-309.

McKenzie, C. R. M., & Mikkelsen, L. A. (2000). The psychological side of Hempel's paradox of confirmation. *Psychonomic Bulletin & Review*, *7*, 360-366.

McKenzie, C. R. M., & Mikkelsen, L. A. (2007). A Bayesian view of covariation assessment. *Cognitive Psychology*, 54, 33–61.

Mirabile, P., & Douven, I. (in press). Abductive conditionals as a test case for inferentialism. *Cognition.*

Mrad, A. B., Delcroix, V., Piechowiak, S., Leicester, P., Mohamed, A. (2015). An explication of uncertain evidence in Bayesian networks: likelihood evidence and probabilistic evidence. *Applied Intelligence*, *43*, 802–824. https://doi.org/10.1007/s10489-015-0678-6

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39-57.

Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 87-109.

Oaksford, M. (2002). Contrast classes and matching bias as explanations of the effects of negation on conditional reasoning. *Thinking & Reasoning*, *8*, 135-151.

Oaksford, M. & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, *101*, 608-631.

Oaksford, M. & Chater, N. (1996). Rational explanation of the selection task. *Psychological Review*, *103*, 381-391.

Oaksford, M., & Chater, N. (2003a). Computational levels and conditional reasoning. *Journal of Experimental Psychology*: *Learning, Memory & Cognition*, *29*, 150-156.

Oaksford, M., & Chater, N. (2003b). Optimal data selection: Revision, review and re-evaluation. *Psychonomic Bulletin & Review*, *10*, 289-318.

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

Oaksford, M., & Chater, N. (2010a). Conditionals and constraint satisfaction: Reconciling mental models and the probabilistic approach? In M. Oaksford & N. Chater (Eds.), *Cognition and conditionals*: *Probability and logic in human thinking* (pp. 309-334). Oxford: Oxford University Press.

Oaksford, M., & Chater, N. (2010b). Causation and conditionals in the cognitive science of human reasoning. *Open Psychology Journal*, *3*, 105-118. (Special issue on *Causal learning beyond causal judgement*, Edited by J. C. Perales, & D. R. Shanks).

Oaksford, M., & Chater, N. (2013). Dynamic inference and everyday conditional reasoning in the new paradigm. *Thinking and Reasoning, 19*, 346-379. doi: 10.1080/13546783.2013.808163

Oaksford, M., & Chater, N. (2016). Probabilities, causation, and logic programming in conditional reasoning: A reply to Stenning and van Lambalgen. *Thinking and Reasoning*, *22*, 336-354. doi: 10.1080/13546783.2016.1139505

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

Oaksford, M., & Chater, N. (2020a). New paradigms in the psychology of reasoning. *Annual Review of Psychology*, *71*, 305-330.

Oaksford, M., & Chater, N. (2020b). Integrating Causal Bayes Nets and inferentialism in conditional inference. In S. Elqayam, I. Douven, J. St.B. T. Evans, & N. Cruz (Eds.). *Logic and uncertainty in the human mind: A tribute to David E. Over*. London: Routledge.

Oaksford, M., Chater, N., & Grainger, B. (1999). Probabilistic effects in data selection. *Thinking and Reasoning*, *5*, 193–243.

Oaksford, M., Chater, N., Grainger, B., & Larkin, J. (1997). Optimal data selection in the reduced array selection task (RAST). *Journal of Experimental Psychology: Learning, Memory & Cognition*, *23*, 441-458.

Oaksford, M., Chater, N., & Larkin, J. (2000). Probabilities and polarity biases in conditional inference. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *26*, 883-899.

Oaksford, M., & Moussakowski, M. (2004). Negations and natural sampling in data selection: Ecological vs. heuristic explanations of matching bias. *Memory & Cognition*, *32*, 570-581.

Oaksford, M., Over, D. E., & Cruz, N. (2019). Paradigms, possibilities and probabilities: Comment on Hinterecker, Knauff, and Johnson-Laird (2016). *Journal of Experimental Psychology*: *Learning, Memory, & Cognition*, *45*, 288-297.

Oaksford, M., & Stenning, K. (1992). Reasoning with conditionals containing negated constituents. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *18,* 835–854.

Oaksford, M., & Wakefield, M. (2003). Data selection and natural sampling: Probabilities do matter. *Memory & Cognition*, *31*, 143-154.

Oberauer, K. (2006). Reasoning with conditionals: A test of formal models of four theories. *Cognitive Psychology*, *53*, 238–283. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1016/j.cogpsych.2006.04.001

Oberauer, K., & Wilhelm, O. (2003). The meaning(s) of conditionals: Conditional probabilities, mental models, and personal utilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 680–693. http://dx.doi.org/10.1037/0278-7393.29.4.680

Oberauer, K., Weidenfeld, A., & Hörnig, R. (2004). Logical reasoning and probabilities: A comprehensive test of Oaksford and Chater (2001). *Psychonomic Bulletin & Review*, *11*, 521–527. https://doi-org.ezproxy.lib.bbk.ac.uk/10.3758/BF03196605

Orenes, I., Beltrán, D., & Santamaría, C. (2014). How negation is understood: Evidence from the visual world paradigm. *Journal of Memory and Language*, *74*, 36–45. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1016/j.jml.2014.04.001

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*, 62–97. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1016/j.cogpsych.2006.05.002

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. San Mateo: Morgan Kaufmann.

Pfeifer, N., & Kleiter, G. D. (2009). Framing human inference by coherence based probability logic. *Journal of Applied Logic*, *7*, 206-217.

Politzer, G., & Baratgin, J. (2016). Deductive schemas with uncertain premises using qualitative probability expressions. *Thinking & Reasoning*, *22*, 78-98.

Pollard, P., & Evans, J. St.B. T. (1983). The effect of experimentally contrived experience on reasoning performance. *Psychological Research*, *45*, 287–301. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1007/BF00308708

Pothos, E. M., & Busemeyer, J. R. (2013). Can quantum probability provide a new direction for cognitive modeling? *Behavioral and Brain Sciences*, *36*, 255–274. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1017/S0140525X12001525

Pothos, E. M., Busemeyer, J. R., Shiffrin, R. M., & Yearsley, J. M. (2017). The rational status of quantum cognition. *Journal of Experimental Psychology: General*, *146*, 968–987. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1037/xge0000312

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rehder, B. (2003a). A causal-model theory of conceptual representation and categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 1141–1159.

Rehder, B. (2003b). Categorization as causal reasoning. *Cognitive Science, 27*, 709–748.

Rehder, B. (2014). Independence and dependence in human causal reasoning. *Cognitive Psychology, 72*, 54–107.

Rehder, B. (2017). Concepts as causal models: Classification. In M. R. Waldmann (Ed.). *The Oxford handbook of causal reasoning* (pp. 346–375). New York, NY: Oxford University Press.

Rehder, B., (2018). Beyond Markov: Accounting for independence violations in causal reasoning, *Cognitive Psychology*, *103*, 42–84

Rips, L. J. (1994). *The psychology of proof: Deductive reasoning in human thinking*. Cambridge, MA: The MIT Press.

Rottman, B., & Hastie, R. (2016). Do people reason rationally about causally related events? Markov violations, weak inferences, and failures of explaining away. *Cognitive Psychology, 87*, 88–134.

Rozenblit, L. & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science*, *26*, 521-562.

Sanborn, A. N., & Chater, N. (2016). Bayesian brains without probabilities. *Trends in Cognitive Sciences*, *20*, 883–893.

Schroyens, W., & Schaeken, W. (2003). A critique of Oaksford, Chater, and Larkin's (2000) conditional probability model of conditional reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 140–149. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1037/0278-7393.29.1.140

Schroyens, W., Schaeken, W., & d'Ydewalle (2001). The processing of negations in conditional reasoning: A meta-analytic case study in mental logic and/or mental models. *Thinking and Reasoning*, *7*, 121–172.

Schroyens, W., Schaeken, W., Fias, W., & d'Ydewalle (2000). Heuristic and analytic processes in propositional reasoning with negatives. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *26*, 1713–1734.

Schroyens, W., Schaeken, W., Verschueren, N., & d'Ydewalle (2000). Conditional reasoning with negations: Implicit and explicit affirmation or denial and the role of contrast classes. *Thinking and Reasoning*, *6*, 221–252.

Schwan, B., & Stern, R. (2017). A causal understanding of when and when not to Jeffrey conditionalize. *Philosophers' Imprint*, *17*, 8.

Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, *1*, 54–87. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1016/0022-2496(64)90017-3

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2019). afex: Analysis of Factorial Experiments. R package version 0.23-0. https://CRAN.R-project.org/package=afex

Singmann, H., Klauer, K. C., & Over, D. E. (2014). New normative standards of conditional reasoning and the dual-source model. *Frontiers in Psychology*, *5*, 316.

Skovgaard-Olsen, N., Collins, P., Krzyżanowska, K., Hahn, U., & Klauer, K. C. (2019). Cancellation, negation, and rejection. *Cognitive Psychology*, *108*, 42–71. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1016/j.cogpsych.2018.11.002

Skovgaard-Olsen, N., Kellen, D., Hahn, U., & Klauer, K. C. (2019). Norm conflicts and conditionals. *Psychological Review*, *126*, 611–633. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1037/rev0000150.supp (Supplemental)

Skovgaard-Olsen, N., Kellen, D., Krahl, H., & Klauer, K. C. (2017). Relevance differently affects the truth, acceptability, and probability evaluations of 'and', 'but', 'therefore', and "if–then." *Thinking & Reasoning*, *23*, 449–482. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1080/13546783.2017.1374306

Skovgaard-Olsen, N., Singmann, H., & Klauer, K. C. (2016). The relevance effect and conditionals. *Cognition*, *150*, 26–36. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1016/j.cognition.2015.12.017

Skovgaard-Olsen, N., Singmann, H., & Klauer, K. C. (2017). Relevance and reason relations. *Cognitive Science*, *41*, 1202–1215. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1111/cogs.12462

Sloman, S. A., & Fernbach, P. (2017). *The knowledge illusion: Why we never think alone*. New York, NY: Riverhead books.

Stahl, C., Klauer, K. C., & Erdfelder, E. (2008). Matching bias in the selection task is not eliminated by explicit negations. *Thinking & Reasoning*, *14*, 281–303. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1080/13546780802116807

Stewart, N., Chater, N., & Brown, G. D. A. (2006). Decision by sampling. *Cognitive Psychology*, *53*, 1–26. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1016/j.cogpsych.2005.10.003

Stewart, N., Chater, N., Stott, H. P., & Reimers, S. (2003). Prospect relativity: How choice options influence decision under risk. *Journal of Experimental Psychology: General*, *132*, 23–46.

Tentori, K., Crupi, V., & Russo, S. (2013). On the determinants of the conjunction fallacy Probability versus inductive confirmation. *Journal of Experimental Psychology: General*, *142*, 235–255. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1037/a0028770

Thompson, V. A., Evans, J. St.B. T., & Campbell, J. I. D. (2013). Matching bias on the selection task: It's fast and feels good. *Thinking & Reasoning*, 19(, 431–452. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1080/13546783.2013.820220

Tsai, T., & Gill, J. (2013). Interactions in generalized linear models: Theoretical issues and an application to personal vote-earning attributes. *Social Sciences*, *2*, 91–113

MTurk (2017). https://docs.aws.amazon.com/AWSMechTurk/latest/AWSMturkAPI/Welcome.html

Vehtari A, Gelman A, Gabry J (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*, 1413-1432. doi: 10.1007/s11222-016-9696-4 (URL: http://doi.org/10.1007/s11222-016-9696-4).

Vehtari, A., Gabry, J., Yao, Y., Gelman, A. (2019). loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models. R package version 2.1.0, <URL: https://CRAN.R-project.org/package=loo>.

Vineberg, S. (2011). Dutch book arguments. In E.N.Zalta (Ed.), *The Stanford Encyclopedia of Philosophy,* Summer 2011 Edition., available at http://plato.stanford.edu/archives/sum2011/entries/dutch-book/(accessedJan-uary4,2014).

Vul, E., Goodman, N., Griffiths, T. L., & Tenenbaum, J. B. (2014). One and done? Optimal decisions from very few samples. *Cognitive Science*, *38*, 599-637.

Ward, W. C., & Jenkins, H. M. (1965). The display of information and the judgement of contingency. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *19*, 231–241. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1037/h0082908

Wason, P. C. (1965). The contexts of plausible denial. *Journal of Verbal Learning & Verbal Behavior*, *4*, 7-11.

Wulff, D. U., Mergenthaler-Canseco, M., & Hertwig, R. (2018). A meta-analytic review of two modes of learning and the description-experience gap. *Psychological Bulletin*, *144*, 140–176. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1037/bul0000115.

Yama, H. (2001). Matching versus optimal data selection in the Wason selection task. *Thinking and Reasoning*, *7*, 295–311.

Yao, Y., Vehtari, A., Simpson, D., & Gelman, A. (2018). Using stacking to average Bayesian predictive distributions (with Discussion). *Bayesian Analysis*, *13*, 917--1007. doi:10.1214/17-BA1091. https://projecteuclid.org/euclid.ba/1516093227

Appendices

Appendix A1

Table A1 shows the best fitting parameter values for the implicit negations data from the studies cited in the section *Modelling the default prior*. We used the **DEoptim** function in R (Ardia, Mullen, Peterson, & Ulrich, 2016) to find the globally optimal cell values of the JPD providing the best fits to the overall frequency of inference endorsements in these studies.

Table A1

The best-fit parameter value for the four rules in the implicit negations paradigm task.

	If p_1 then q_1				*If p_1 then $\neg q_1$*			
	q_1	q_2	q_3	Total	q_1	q_2	q_3	Total
p_1	0.568	0.000	0.015	0.583	0.028	0.224	0.102	0.354
p_2	0.163	0.084	0.011	0.258	0.049	0.136	0.159	0.344
p_3	0.061	0.089	0.007	0.157	0.075	0.011	0.216	0.302
Total	0.792	0.173	0.033	1.000	0.152	0.371	0.477	1.000
	If $\neg p_1$ then q_1				*If $\neg p_1$ then $\neg q_1$*			
p_1	0.106	0.041	0.146	0.293	0.260	0.052	0.219	0.531
p_2	0.260	0.026	0.096	0.382	0.170	0.094	0.063	0.327
p_3	0.132	0.005	0.189	0.326	0.017	0.080	0.045	0.142
Total	0.498	0.072	0.431	1.000	0.447	0.226	0.327	1.000

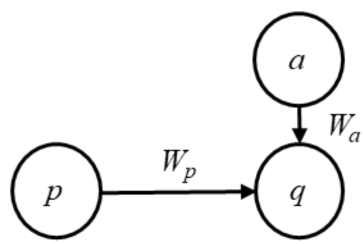
Supplementary Online Material

Causal Bayes nets.

We have argued that people mentally represent conditionals in a similar way to causal Bayes nets (Ali, Chater, & Oaksford, 2011; Ali, Schlottman, Shaw, Chater, & Oaksford, 2010; Chater & Oaksford, 2006; Oaksford & Chater, 2010b, 2013, 2016, 2017). Figure S1 shows how we can implement the JPD Pr_1 in Table 1 in a Causal Bayes net where the weights on the directed links correspond to causal powers, W_p (Cheng, 1997). In this network *travelling to Manchester* is treated as the cause of Johnny *taking the train*, although there may be alternative causes, *a*, of him travelling by train.

Figure S1

Causal Bayes Net implementing the JPD Pr₁ in Table 1 interpreted causally



$$W_p = \frac{\Pr(q|p) - \Pr(q|\neg p)}{1 - \Pr(q|\neg p)} = .7$$

$$W_a = \Pr(q|\neg p) = .167,\ \Pr(p) = .4$$

In this causal Bayes net, the cause (p) and its alternative (a) are combined using the noisy-OR integration rule (Pearl, 1988):

$$\Pr(q = 1|p = 1) = 1 - (1 - W_a)(1 - W_p)^{ind(p)} \qquad \text{(Eq. S1)}$$

Where $ind(p) = 1$ when the cause is present ($p = 1$) and $ind(p) = 0$ when the cause is absent ($p = 0$).