



BIROn - Birkbeck Institutional Research Online

Vance, James and Oaksford, Mike (2021) Explaining the implicit negations effect in conditional inference: experience, probabilities and contrast Sets. *Journal of Experimental Psychology: General* 150 (2), pp. 354-384. ISSN 0096-3445.

Downloaded from: <https://eprints.bbk.ac.uk/id/eprint/32295/>

Usage Guidelines:

Please refer to usage guidelines at <https://eprints.bbk.ac.uk/policies.html>
contact lib-eprints@bbk.ac.uk.

or alternatively

1 © 2020, American Psychological Association. This paper is not the copy
2 of record and may not exactly replicate the final, authoritative version of
3 the article. Please do not copy or cite without authors' permission. The
4 final article will be available, upon publication, via its DOI:

5 [10.1037/xge0000954](https://doi.org/10.1037/xge0000954)

6

7 Explaining the Implicit Negations Effect in Conditional Inference:

8

Probabilities, Experience, and Contrast Sets

9

James Vance & Mike Oaksford

10

Birkbeck College, University of London

11

12

13 Author Note

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

20
21
22
23
24
25
26
27
28
29
30
31
32
33
34

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.

35 © 2020, American Psychological Association. This paper is not the copy of record and may not
36 exactly replicate the final, authoritative version of the article. Please do not copy or cite without
37 authors' permission. The final article will be available, upon publication, via its DOI:
38 10.1037/xge0000954
39

40 Explaining the Implicit Negations Effect in Conditional Inference: Probabilities, Experience, and
41 Contrast Sets
42

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

47
48 The psychology of judgement, decision making, causal, and deductive reasoning reveals many
49 apparent biases. Biases are systematic deviations from the predictions of a normative theory of
50 how people should respond on a task. Explaining these biases is a major industry in cognitive
51 psychology/science that has driven many important theoretical developments. Common patterns
52 of explanation are that the wrong normative theory has been applied to a task (Oaksford &
53 Chater, 1994, 2007; Pothos & Busemeyer, 2013; Pothos, Busemeyer, Shiffrin, & Yearsley, 2017);
54 that people are responding to a different question that has an equally normative answer (Griffiths,
55 & Tenenbaum, 2005; Tentori, Crupi, & Russo, 2013); the information was not presented in an
56 understandable format (Gigerenzer & Hoffrage, 1995; Hogarth, & Soyer, 2011; Jarvstad, Hahn,
57 Rushton, & Warren, 2013; Wulff, Mergenthaler-Canseco, & Hertwig, 2018); we need to take
58 account of noise (Costello & Watts, 2014; Costello, Watts, & Fisher, 2018); or that the
59 mind/brain approximates probabilities by sampling (Dasgupta, Schulz, & Gershman, 2017;
60 Hattori, 2016; Sanborn & Chater, 2016; Stewart, Chater, & Brown, 2006), an approach aligned
61 with the classical strategy in the psychology of deductive reasoning of explaining biases at the

62 algorithmic not computational level (Johnson-Laird, 1983; Rips, 1994). Most of these
63 explanations explain away biases while retaining the normative standard of rationality given by
64 classical binary logic (mental logic/mental models) or Bayesian probability theory.¹ That we are
65 beginning to understand the sources of bias in judgement and decision making also resolves a
66 paradox. Explaining biases in the psychology of deductive reasoning, like confirmation bias, has
67 invoked Bayesian probability theory as a normative standard (Oaksford & Chater, 1994, 2007,
68 2020a). Yet, paradoxically, Bayesian reasoning in judgement and decision-making had seemed
69 equally biased. It also opens up the possibility that the way that biases have been explained away
70 in judgement and decision-making may also apply to the psychology of deductive reasoning.

71 In this paper, we investigate a key outstanding problem in the psychology of conditional
72 inference, that is, reasoning with *if p then q* in English, where *p* is the antecedent and *q* the
73 consequent. Polarity biases occur when negations (“not”) are varied in conditionals (Evans,
74 1972, 1998; Evans & Lynch, 1973; Oaksford, 2002; Oaksford & Chater, 1994; Oaksford &
75 Stenning, 1992; Oaksford & Mousakowski, 2004; Schroyens, Schaeken, Fias, & d’Ydewalle,
76 2000; Schroyens, Schaeken, & d’Ydewalle, 2001; Schroyens, Schaeken, Verschueren, &
77 d’Ydewalle, 2000; Yama, 2001). As our opening quotation from Horn (1989) indicates,
78 negations are a defining feature of human linguistic communication. The Aristotelean foundation
79 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).

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

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

100 The implicit negation effect occurs not only for MT but also for the logical fallacies of
 101 *denying the antecedent* (DA: *if p , then q , $\neg p$, therefore $\neg q$*) and *affirming the consequent* (AC: *if*
 102 *p , then q , q , therefore p), and for the other logically valid inference rule of *modus ponens* (MP: *if**

103 *p*, then *q*, *p*, therefore *q*). For example, the AC inference on *if not A, then not 2* using an explicit
 104 negation, *not 2*, produces 61% endorsements of the conclusion, *not A*. In contrast, using an
 105 implicit negation, *7*, causes this to fall to 11% (Evans & Handley, 1999, Expt. 3). Although
 106 implicit negations remove negative conclusion bias, they do not lead to logical performance.
 107 They reduce conclusion endorsements as much for logically valid inferences (MP, MT) as for
 108 logical fallacies (DA, AC).

109 Explanations of this effect may discriminate between the Bayesian new paradigm
 110 approach (Oaksford, 2002; Oaksford, Chater, & Larkin, 2000; Oaksford & Chater, 2003, 2007,
 111 2020a), heuristic approaches (Evans, 1998; Evans et al., 1996; Evans & Handley, 1999), and
 112 mental models theory (Johnson-Laird & Byrne, 2002; Khemlani, Orenes, & Johnson-Laird,
 113 2012), but the critical tests have never been conducted.² Our experiments attempt to provide
 114 these tests. They used probability manipulations shown in decision making to improve
 115 participants' understanding of a task and to lead to better fits to the data (Jarvstad et al., 2013;
 116 Wulf, et al., 2018). We used short experiential learning phases and asked participants for their
 117 subjective estimates of the learned probabilities that we used to predict the results on the
 118 inference task. This is the first time that discrete experiential learning has been used to
 119 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).

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

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

130

131 **Probabilities and Contrast Sets**

132 The new Bayesian paradigm in human reasoning is a broad church (Oaksford & Chater, 2020a).
 133 However, there are several assumptions common to these approaches. First, the conditional is not
 134 a binary truth functional operator, as in the standard logic, that licenses the validity of MP and
 135 MT and not of AC and DA. Second, the probability of a conditional is the conditional
 136 probability, $\Pr(\textit{if } p \textit{ then } q) = \Pr(q|p)$.³ This assumption is called “the Equation” (Edgington,
 137 1995). Third, probabilities are subjective and relate to individuals’ degrees of belief. Finally,
 138 conditional probabilities are suppositional and determined by the Ramsey test: suppose p is true,
 139 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 if p is true and q is false, and true otherwise. Consequently, $\Pr(\textit{if } p \textit{ then } q) = \Pr(p, q) + \Pr(\neg p, q) + \Pr(\neg p, \neg q)$, an assignment that is very rarely observed empirically.

140 There are a variety of sophisticated probabilistic approaches to conditional inference, for
 141 example, probability logic (Cruz, Baratgin, Oaksford, & Over, 2015; Evans, Thompson, & Over,
 142 2015; Pfeifer & Kleiter, 2009; Politzer & Baratgin, 2016; Singmann, Klauer, & Over, 2014),
 143 belief revision (Eva & Hartmann, 2018; Oaksford & Chater, 2007, 2010b, 2013), and Bayes nets
 144 (Ali, Chater, & Oaksford, 2011; Chater & Oaksford, 2006; Fernbach & Erb, 2013; Oaksford &
 145 Chater, 2010b, 2013, 2017). We will discuss these in the sequel. For now, as a first
 146 approximation, we assume that the probability of a conclusion of an inference is its conditional
 147 probability given the categorical premise calculated over a joint probability distribution (JPD)
 148 (Anderson, 1995; Oaksford et al., 2000).⁴ We can then derive our predictions by considering two
 149 JPDs one without (Table 1) and one with contrast sets (Table 2).

150
 151 Table 1
 152 *Learning a new distribution*
 153

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

154

155

156 **Contradictory Negation**

157 Suppose your initial beliefs about Johnny’s travelling habits are captured by the JPD Pr_0 in Table
 158 1. In this table, p and $\neg p$ are contradictories, and are treated with “absolute symmetry” (Horn,
 159 1989, p. xiii). If one of these propositions is true the other is false, but finding out that Johnny
 160 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.

161 In Pr_0 , you are reasonably confident that *if he travels to Manchester (p), he takes the train*
 162 *(q)*. Your degree of belief in the conditional is the relevant conditional probability computed over
 163 this JPD, $Pr_0(q|p) = .75$. However, you are maximally uncertain about whether he takes the train
 164 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
 165 just less than half of his journeys are to Manchester ($Pr_0(p) = .4$). Now suppose either that you
 166 learn (1) from experience or a reasonably reliable informant.

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

168 We assume that the result of learning or hearing (1) from a reliable source, leads you to revise
 169 your beliefs about Johnny's travelling habits to the JPD Pr_1 in Table 1, in which $Pr_1(\neg q|\neg p) =$
 170 $Pr_1(\neg p, \neg q)/Pr_1(\neg p) = .5/.6 = .833$.⁵ In our experiments, we provide people with relevant
 171 experience to revise their beliefs from Pr_0 to Pr_1 , where Pr_1 implements manipulations designed
 172 to test our account of the implicit negations effect. In the sequel, we fit the model to previous
 173 data to estimate people's default prior beliefs, Pr_0 .

174 Suppose you then learn that, on a particular journey, Johnny did not take the train. With
 175 what probability should you now believe that he did not go to Manchester? We treat this query as
 176 the probabilistic equivalent of an AC inference having learned (1), and with *Johnny did not take*
 177 *the train* as the categorical premise. As we have said, for now, we treat the probability of the
 178 conclusion of an inference as the conditional probability of the conclusion given the categorical
 179 premise calculated over the JPD Pr_1 in Table 1 (Anderson, 1995; Oaksford et al., 2000). So for
 180 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.

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

183

184 **Contrary Negation: Contrast Sets**

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

193 The philosophical and linguistic depiction of negation as otherness—negated statements
 194 make a positive reference to something other than the negated proposition—can be traced back
 195 to Plato and to Aristotle's account of contrary negation (Horn, 1989). The variety of ways in
 196 which people can use and express negation in natural languages (Horn, 1989) means that
 197 identifying contrast sets could not be their sole function. However, they can explain polarity
 198 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).

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

201
 202 Table 2.
 203 *A joint probability distribution for implicit negations.*
 204

| | 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) |

205
 206 *Note. Frequencies of occurrence in the learning trials in Experiment 1 are shown in brackets.*
 207

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

218 As for Pr_1 in Table 1, knowing the distribution in Table 2 may lead someone to accept (1).
 219 On being told *Johnny did not travel to Manchester*, they should then still endorse the conclusion

220 of the MP inference on (1), *he did not take the train*, quite strongly, because in the JPD in Table
 221 2, $\Pr(\neg q|\neg p) = (\Pr(p_2, q_2) + \Pr(p_2, q_3) + \Pr(p_3, q_2) + \Pr(p_3, q_3))/(\Pr(p_2) + \Pr(p_3)) = .5/.6 = .833$.
 222 However, if told that *Johnny travelled to Paris*, then the probability that *he did not take the train*,
 223 $\Pr(\neg q|p_2) = (\Pr(p_2, q_2) + \Pr(p_2, q_3))/\Pr(p_2) = .06/.16 = .375$, which predicts much lower
 224 endorsement of MP. We would expect an implicit negations effect.

225 All other theories of the implicit negation effect argue that it arises solely from using an
 226 implicit negation, regardless of probabilistic structure. However, Table 2 suggests that we should
 227 be able remove the effect even when using an implicit negation in the categorical premise. If q_3 ,
 228 *he travelled by plane*, is used to affirm the consequent of (1), $\neg q_1$, then Table 2 does *not* predict
 229 an implicit negation effect for AC for this conditional. In this JPD, $\Pr(\neg p|\neg q) = (\Pr(p_2, q_2) +$
 230 $\Pr(p_2, q_3) + \Pr(p_3, q_2) + \Pr(p_3, q_3))/(\Pr(q_2) + \Pr(q_3)) = .833$, and $\Pr(\neg p|q_3) = (\Pr(p_2, q_3) + \Pr(p_3,$
 231 $q_3))/\Pr(q_3) = .24/.30 = .80$. Consequently, whether using an explicit negation (AC-Not) or an
 232 implicit negation drawn from the contrast set (AC-Con), people should endorse AC almost
 233 equally often. This prediction, that the implicit negations effect depends on probabilistic
 234 structure, discriminates the probabilistic contrast set theory from all other theories.

235

236 **Experience: Manipulating Probabilities**

237 Testing these predictions requires manipulating probabilities. Reasoning researchers have
 238 manipulated probabilities in many ways, using pre-tested content (Oaksford, et al., 2000;
 239 Oaksford, Chater, & Grainger, 1999), frequency formats (Gigerenzer & Hoffrage, 1995)
 240 combined with concrete visualizations (stacks of cards) (Oaksford, et al., 1997, 1999),
 241 contingency tables, or “probabilistic truth tables” (Evans, Handley, & Over, 2003; Oberauer &
 242 Wlihelm, 2003), as in causal judgement (Ward & Jenkins, 1965), a procedure that has also been

243 reversed so participants construct the contingency table given a conditional (Oaksford &
244 Mousakowski, 2004; Oaksford & Wakefield, 2003; Oberauer, 2006; Over, Hadjichristidis,
245 Evans, Handley, & Sloman, 2007), and sequential tasks where trial frequency reflects the
246 probabilities (Fugard, Pfeifer, Mayrhofer, & Kleiter, 2011; Oaksford & Mousakowski, 2004;
247 Oaksford & Wakefield, 2003), and where learning effects are observed (for critiques, see Jubin
248 & Barrouillet, 2019; Oberauer, Weidenfeld, & Hörnig, 2004). In these experiments, we used
249 experiential learning of probabilities, which leads to improved performance in judgment and
250 decision-making, and which has not used before in reasoning research.

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

262 No other theory of the implicit negations effect predicts that learning about
263 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).

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

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

278 In conditional inference, the most relevant distribution from which we could provide
279 samples are the conditional probabilities that correspond to people's predicted degree of belief in
280 the conclusion of the inferences MP, DA, AC, and MT (see Table 3 below). However, as for
281 probability judgement, providing such samples is rather too close to giving participants the
282 probabilistically correct answer (Hawkins et al., 2015). Although we wanted to exploit the
283 potential benefits of trial-by-trial sampling, we also wanted to assess people's ability to
284 extrapolate from information that they might experience in the real world. Therefore, we used
285 experiential trial-by-trial learning of the JPD in Table 2, to get participants to revise their default

286 prior beliefs, Pr_0 , to a new distribution, Pr_1 , which implements the focused manipulations that
287 test our account of the implicit negation effect.

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

298

299

Experiment 1: MP Manipulation

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

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

312 In this experiment, participants drew inferences before and after the learning phase. We
313 presented single event probability descriptions (e.g., 0.8 or 80%) before the pre-learning
314 inference task. In this phase, we predicted that we would observe the default implicit negations
315 effect, based on the default prior (Pr_0), for these materials. Previous evidence showed an implicit
316 negation effect for this conditional (*if* $\neg p$, *then* $\neg q$) for both MP (MP-Con [44%] < MP-Not
317 [89%]) and AC (AC-Con [11%] < AC-Not [61%]) (Evans & Handley, 1999, Experiment 3).
318 Moreover, in a meta-analysis of previous results, the sample size weighted mean decrease in
319 percentage endorsements for explicit vs implicit negations was 42% for MP, and 57% for AC
320 (Evans & Handley, 1999; Schroyens et al., 2000). Consequently, in this experiment we also
321 tested Hypothesis 2:

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

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

329 In decision making, using participants' subjective estimates of learned probabilities, also
330 provides better fits to the data than objective values (Jarvstad et al, 2013). Consequently, in these

331 experiments, on completing the inference task, we asked participants to perform a probability
332 verification task where they reconstructed the JPD in Table 2. This procedure allowed us to
333 check how well participants had learned this distribution by computing the correlation with the
334 objective values. Splitting participants in to high and low correlation groups will also allow us to
335 see how well the probabilities are learned affects inference. We also used these joint probabilities
336 to calculate the relevant conditional probabilities for each inference. We could then test how well
337 these subjective calculated conditional probabilities predicted inference task performance, which
338 leads to our third hypothesis:

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

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

352 *Hypothesis 4.* Subjective probability will directly predict confidence, or

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

355

356 **Analysis Strategy**

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

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

365 **Comparing means.** We used the R packages **tidybayes** (Kay, 2019) and **emmeans**
 366 (Lenth, 2019), to generate samples for each marginal mean. When comparing means, we
 367 assumed a region of practical equivalence (ROPE, Kruschke, 2011) of $0 \pm 0.1 \times \text{SD}$ of the
 368 differences, and report the proportion of the distribution of differences falling outside the ROPE.
 369 This procedure avoids the unrealistic assumption of a point null hypothesis. We report this
 370 statistic, where the proportion is p , as “ $p \notin \text{ROPE}$ ”.⁸ We also computed Cohen’s d for each
 371 comparison. For all means and differences, we report the 95% highest density interval (HDI) in
 372 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 $0.1 \times \text{SD}$.

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

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

387

388 **Method**

389 **Participants.** Participants were recruited via Amazon Mechanical Turk (2017). Sample
390 size was determined both classically (Chow, Shao, & Wang, 2008) and by Bayesian estimation
391 using the **proddiff.mblacc** function from the **SampleSizeProportions** package in R (Joseph, du
392 Berger, & Belisle, 1997). Previous data from Evans and Handley (1999) for the AC inference on
393 *if $\neg p$ then $\neg q$* was used to estimate effects size. Maintaining a 5% chance of a Type 1 error and a
394 20% chance of Type 2 error, led to very different required sample sizes; classical: 22 (11 in each
395 group), Bayesian: 244 (122 in each group). One of our key predictions is an interaction, and

396 reliably estimating interactions requires sixteen times more data than main effects (Gelman,
397 2018). Consequently, recruitment aimed for a sample size of between 250 and 300.

398 All participants who completed the experiment received a small payment (between
399 US\$0.50 and US\$1.00). Some responses were excluded because they may have been duplicates,
400 either sharing an MTurk ID or an IP address. After exclusions, the sample size was 272. 52%
401 were female and the sample was aged between 18 and 75 with a median age of 31 years (mean =
402 34.34, SD = 11.94). English was the first language of 97% of participants.

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

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

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

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

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

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

419 “Please select the option below that best describes what she should conclude about the
 420 next animal.”

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

| | | |
|-----|-------------------------------------|-------------------------------------|
| 423 | MP and DA alternatives: | AC and MT alternatives: |
| 424 | <i>That the animal is not black</i> | <i>That the animal is not a cat</i> |
| 425 | <i>That the animal is black</i> | <i>That the animal is a cat</i> |

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

428
 429 Table 3

430 *The Probabilities of Drawing Each Inference in Experiment 1*

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

431
 432 *Note: Inf. = Inference*
 433

434 The experiment also included a learning phase with 50 trials. Each trial consisted of a
 435 photograph of one of the 50 animal/colour pairings shown in Table 1. Each photograph showed
 436 only the animal against a white background. Each of the 50 photographs was unique. So, for
 437 example, participants would see 15 different black cats, and so on. The photographs were

438 cropped and re-sized so that they were the same size and fitted on to a single screen at typical
439 resolution for online presentation. The pictures were presented in random order. To try and
440 ensure that participants attended to the stimuli, on each trial, the participant had to answer two
441 questions with three response options each: *What type of animal is this? (Dog, Cat, Rabbit)*, and
442 *What colour best describes this animal? (Black, White, Brown)*.

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

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

458 The participants were then given instructions for the learning phase, as in the *Design and*
459 *Materials* section, where they were told they would see a sample of the animals that the vet sees
460 in the surgery. Participants then performed the *post-learning* phase inference task, this time with

461 no information about the proportion of animals. Finally, participants were presented with a
462 probability verification task to check how accurately they could reconstruct the probability
463 distribution in Table 2. Each participants' subjective conditional probabilities of drawing each
464 inference could then be calculated. This task consisted of nine response boxes in a three by three
465 grid labelled animal type (cat, dog, rabbit) on one axis and colour (black, white, brown) on the
466 other, as in Table 2. Participants were instructed to enter how many of the next 100 animals that
467 the vet would see would be in each category (a similar procedure was used in Oaksford &
468 Wakefield, 2003). If participants attempted to proceed without their responses summing to 100,
469 they were returned to this page with an instruction to make sure their responses did add up to 100
470 and were provided with the total value they initially entered for guidance.

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

474

475 **Results and Discussion**

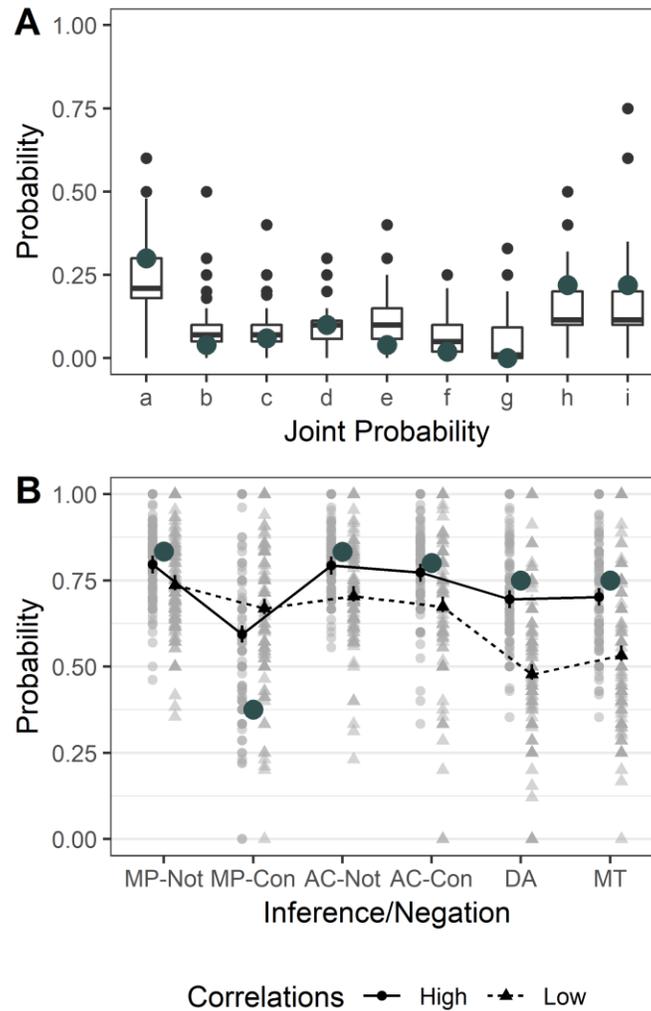
476 **Attention test.** The attention test in the learning task involved naming the animal and
477 colour on each trial. With fifty trials, each participant could make up to 100 errors. The mean
478 error rate was less than 1% (.70, SD = 2.26). Only 37 participants (13.6%) made more than 1
479 error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

482

483

484

485 Figure 1
 486 *Joint Probabilities and Calculated Conditional Probabilities from the Probability Verification*
 487 *Task in Experiment 1*



488
 489 *Note. A. Box-plots for the verification judgements for all cells of Table 1. B. Mean calculated*
 490 *conditional probabilities for each inference based on the estimates shown in panel A split by*
 491 *correlation with the objective values, error bars = 95% HDI; model: $Cond \sim InfNeg * Corr$. In*
 492 *both panels, the large dark grey points indicate the objective probabilities based on Table 1.*

493
 494 **Probability verification task.** We first report the results of the probability verification
 495 task. Figure 1A shows the box-plots for each cell in Table 2 and the objective values for each
 496 cell. We used the standard letter labelling of cells in a contingency table used in causal learning

497 (Hattori & Oaksford, 2007). Errors for low probabilities can only push in one direction and all
498 cell values must sum to 1. Therefore, unsurprisingly, lower objective values tended to be
499 overestimated and higher values underestimated. The mean correlation between each
500 participant's estimates and the objective values was $r(7) = .59$ ($SD = .33$). We split participants
501 into high and low correlation groups (*Corr*); high correlation (\geq median): mean $r(7) = .81$ (SD
502 $= .11$, $N = 148$), and low correlation ($<$ median): mean $r(7) = .32$ ($SD = .30$, $N = 124$). By this
503 measure, there was a large group of participants who showed a good understanding of the
504 underlying probabilities, but also a group who did not, sometimes showing negative correlations
505 with the objective values.

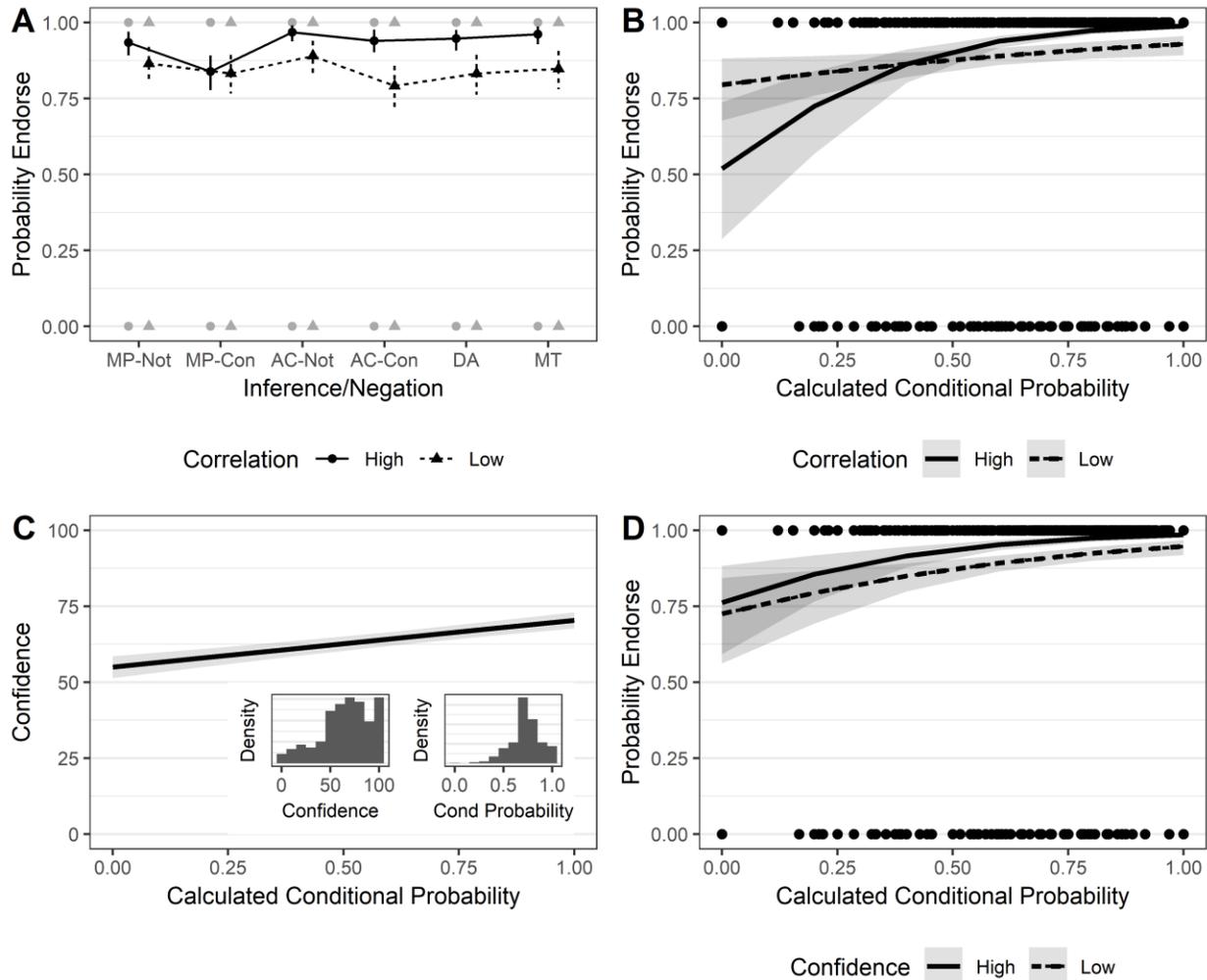
506 We then used the estimated values from the probability verification task to compute the
507 conditional probabilities (*Cond*) for each inference. There were occasional missing data points
508 because of problems of division by zero. To maintain the coherence of the computed conditional
509 probabilities,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

514 Figure 1B shows the estimated marginal means of the posterior distribution (see figure
515 caption for the model). For the high correlation group, MP-Con (mean = .59, 95% HDI =
516 [.57, .62]) was lower than MP-Not (mean = .80 [.78, .82], $\bar{d} = 8.20$ [5.88, 10.47], $1.0 \notin$ ROPE).
517 Exactly the same pattern of differences was observed between MP-Con and the remaining four
518 inferences, AC-Not (mean = .79 [.77, .82], $\bar{d} = 14.90$ [12.65, 17.31]), AC-Con (mean = .77
519 [.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

520 (mean = .70 [.68, .72], $\bar{d} = 8.20 [5.88, 10.48]$). For all comparisons, $1.0 \notin$ ROPE. There were no
 521 differences between MP-Not, AC-Not or AC-Con ($< .93 \notin$ ROPE for all comparisons). For the
 522 AC-Not vs AC-Con comparison 0 was a credible value for the effect size, $\bar{d} = 1.55 [-.77, 3.77]$.
 523 However, all these inferences differed from DA and MT ($1.0 \notin$ ROPE for all comparisons),
 524 although DA and MT did not differ from each other ($.61 \notin$ ROPE). Although the differences
 525 were smaller, the same basic pattern occurred for MP-Not, MP-Con, AC-Not, and AC-Con for
 526 the low correlation group. However, DA (mean = .48 [.45, .50]) and MT (mean = .53 [.51, .55])
 527 were much lower in the low correlation group than the high correlation group ($1.0 \notin$ ROPE for
 528 both comparisons). In summary, for the high correlation group, the calculated conditional
 529 probabilities based on the verification task produced the predicted manipulation such that
 530 $\Pr(\neg q_1 | \neg p_1)$ (MP-Not) $>$ $\Pr(\neg q_1 | p_2)$ (MP-Con), and $\Pr(\neg p_1 | \neg q_1)$ (AC-Not) \approx $\Pr(\neg p_1 | q_3)$ (AC-Con).

531 **Inference Tasks.** We first looked at the results for the pre-learning inference task with
 532 inference (AC, MP) and negation (Not, Con) as categorical predictors. The effect for AC was
 533 larger than the effect for MP. AC-Con (mean = .82 [.77, .86]) was lower than AC-Not (mean
 534 = .87 [.83, .91]) but zero was still a credible value for the difference ($\bar{d} = 2.14 [-.51, 5.06]$, $.94 \notin$
 535 ROPE) but only marginally. In contrast, although MP-Con (mean = .83 [.79, .88]) was lower
 536 than MP-Not (mean = .86 [.82, .90]) zero was a credible value for the difference ($\bar{d} = 1.3 [-1.31,$
 537 $4.13]$, $.80 \notin$ ROPE). No differences were observed between any of the other inferences (0 was a
 538 credible value for all differences and $< .92 \notin$ ROPE for all comparisons). The results of the pre-
 539 learning inference task were consistent with default expectations derived from previous research
 540 where the implicit negation effect is larger for AC than MP, thereby providing some support for
 541 Hypothesis 2. It means that the learning task based on Table 1 must overcome this default prior
 542 to reveal the effects predicted by Hypothesis 1.

543 Figure 2
 544 *The Results of the Post-Learning Inference Phase in Experiment 1*



545
 546 *Notes. A The probability of endorsing each inference for the high and low correlation groups,*
 547 *error bars = 95% HDI; B. The probability of endorsing an inference predicted by the calculated*
 548 *conditional probability for the high and low correlation groups; C. The relationship between*
 549 *calculated conditional probability and confidence for the high correlation group showing density*
 550 *plots for each variable; D. The probability of endorsing an inference predicted by the calculated*
 551 *conditional probability for the high correlation group with high and low confidence.*
 552

553 We first fitted a model to the post learning phase inference task, using inference/negation
 554 and correlation as categorical predictors. The estimated marginal means are shown in Figure 2A.
 555 We then looked at the interaction between inference (*Inf*: MP and AC) and negation (*Neg*: Not,

556 Con) for the high correlation group. We compared two models, one which included the
 557 interaction (M1), and one with only the main effects (M2) (see, Table 4 Note). $\Delta elpd$ and the
 558 Bayesian stacking weights converged on identifying M2 as the best model. It provides the most
 559 efficient compression of the data by minimizing the loss of information using the fewest
 560 parameters. This result suggests that we have failed to observe the predicted interaction.

561 However, $\Delta elpd$ indicates that there was only a small difference between models. M2 is
 562 weighted more heavily because it is simpler, having fewer parameters. Moreover, estimating
 563 interactions requires sixteen times more data than main effects (Gelman, 2018), as we noted in
 564 the *Participants* section. The simple effects were as predicted. MP-Con (mean = .84 [.79, .88])
 565 was lower than MP-Not (mean = .93 [.89, .97]) ($\bar{d} = 3.71 [.63, 4.46]$, .99 \notin ROPE) and AC-
 566 Con (mean = .94 [.90, .98]) ($\bar{d} = 4.03 [1.22, 6.75]$, .99 \notin ROPE). However, zero was a credible
 567 value for the difference between AC-Not (mean = .97 [.94, .99]) and AC-Con ($\bar{d} = 1.57 [-1.30,$
 568 $4.24]$, .85 \notin ROPE) and MP-Not ($\bar{d} = 1.86 [-.91, 4.75]$, .89 \notin ROPE).

569
 570 Table 4

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

| | <i>LOOIC</i> | <i>SE</i> | <i>k</i> | $\Delta LOOIC$ | $\Delta elpd$ | Δse | <i>Weight</i> |
|----|--------------|-----------|----------|----------------|---------------|-------------|---------------|
| M1 | 324.1 | 32.3 | 4.1 | 1.8 | .9 | .5 | 0 |
| M2 | 322.3 | 32.0 | 3.0 | 0 | 0 | 0 | 1.0 |

572
 573 *Note.* M1: $Endorse \sim Inf * Neg$, M2: $Endorse \sim Inf + Neg$. Estimated number of parameters (*k*),
 574 the difference ($\Delta LOOIC$), the difference in expected log posterior predictive density ($\Delta elpd$) and
 575 its standard error (Δse), and the Bayesian stacking weights (*LOOIC-weight*).
 576

577 There was only one difference for the low correlation group. AC-Con (mean = .79
578 [.72, .86]) was lower than AC-Not (mean = .89 [.83, .94]) ($\bar{d} = 2.15$ [.13, 4.02], $.98 \notin \text{ROPE}$).
579 This effect is consistent with the default prior effect we derived from previous results and the
580 results of the pre-learning inference task. It suggests that even though most participants attended
581 to the learning task, the low correlation group did not learn from it and reverted to the default
582 prior.

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

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

599 a foil model (M4), which included just the intercepts to test that including calculated conditional
 600 probability provided more accurate predictions.

601
 602 Table 5.

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

| | <i>LOOIC</i> | <i>SE</i> | <i>k</i> | Δ <i>LOOIC</i> | Δ <i>elpd</i> | Δ <i>se</i> | <i>Weight</i> |
|----|--------------|-----------|----------|-----------------------|----------------------|--------------------|---------------|
| 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 |

605

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

610

611 Table 5 shows the results of the model comparison. The stacking weights and Δ elpd
 612 converged on identifying M3 as the best model. One could argue that M3 provides the better fit
 613 because it contains more parameters (*k*). However, Bayesian indices of fit, like LOOIC and BIC,
 614 heavily penalize model complexity (many parameters), and far more than conventional fit
 615 indices, like AIC⁹. Consequently, that M3 still provides a much better fit is impressive.
 616 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.

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

620 Figure 2B shows the relation between calculated conditional probability and endorsement
621 rates for the high and low correlation groups for M3. Interpreting slopes and interactions is
622 problematic in generalized linear models (Tsai & Gill, 2013). Parameters are estimated after a
623 non-linear logit (i.e., log-odds) transformation of the model. Describing the effects is most
624 interpretable by transforming the dependent variable to odds. The slope for the high correlation
625 group was 129.86 [5.25, 393.63] ($b > 0$, $.97 \notin \text{ROPE}$), that is, a .1 increase in calculated
626 conditional probability increases the odds that an inference will be endorsed by 13. For the low
627 correlation group, the slope was 4.02 [.75, 9.02] ($b > 0$, $1.0 \notin \text{ROPE}$), that is, a .1 increase in
628 calculated conditional probability increases the odds by .4. The intercept for the high correlation
629 group was 1.29 [.21, 2.94], indicating that when the calculated conditional probability was zero,
630 an inference was still marginally more likely to be endorsed than rejected. For the low
631 correlation group the intercept was 4.74 [.41, 12.28]. Intercepts did not differ between groups (\bar{d}
632 = -1.19 [-4.32, 1.22], $.78 \notin \text{ROPE}$), but the slope for the high correlation group was steeper than
633 for the low ($\bar{d} = 1.11$ [.002, 3.44], $.95 \notin \text{ROPE}$).

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

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

642 **Confidence.** We next assessed the relationship between confidence and the calculated
643 conditional probabilities using the model $Confidence \sim Cond + (1|Participant)$. Figure 2C shows
644 that they are linearly related. The population slope was 15.38 [10.22, 20.15] ($b > 0$, $1.0 \notin ROPE$)
645 indicating that a 0.1 increase in conditional probability lead to 1.54 [1.5, 3.1] point rise in
646 confidence. Both distributions were skewed to the high end of the scale (see subplots in Figure
647 2C), and they had median values at the same point (conditional probability: .69; confidence: 69).
648 Consistent with this correlation, Figure 2D shows that the median split on confidence ($ConfSplit$)
649 produced a slightly higher intercept when confidence was high without a change in slope (model:
650 $Endorse \sim Cond*ConfSplit + (1|Participant)$). However, zero was a credible value for the
651 differences between high and low response groups for both the slope and the intercept. These
652 results were not consistent with confidence moderating the effect of conditional probability on
653 endorsements. These results, therefore, confirm Hypothesis 4, but disconfirm Hypothesis 4'.

654 **Possible criticisms.** Before summarising, we consider two possible criticisms of this
655 experiment. First, the 2AFC response mode may result in more polarized results, perhaps
656 favouring a probabilistic explanation. Response mode can alter response patterns in conditional
657 inference, but not by very much (Evans, Clibbens, & Rood, 1995; Evans & Handley, 1999;
658 Oaksford & Chater, 2010a; Schroyens, Schaeken, & d'Ydewalle, 2001). The 2AFC procedure is
659 similar to evaluation tasks where participants see the valid conclusion and its negation separately
660 and are asked for an endorse decision (Marcus & Rips, 1979; Oaksford, et al., 2000). The current
661 procedure combines these separate choices (which, in the aggregate, sum to 1, see Oaksford, et

662 al., 2000), into a single decision, and provides no reason to expect endorsement decisions to
663 diverge from previously used response modes.

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

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

685 inference endorsement (H4). These results are not consistent with other theories of the implicit
 686 negation effect, which all predict an implicit negation effect for both MP and AC.

687

688 **Experiment 2: MP and AC Manipulations**

689 Experiment 1 had some limitations. First, the effects, although statistically reliable with good
 690 effect sizes, were not of the same magnitude observed in the literature on implicit negations.

691 Moreover, they only occurred for the high correlation group. The low correlation group

692 continued to show the default effect also seen in the pre-learning inference task. Second,

693 although the simple effects were all in the predicted direction, we did not observe the predicted

694 interaction. Third, the distribution of calculated conditional probabilities was skewed toward the

695 upper end of the scale. Such an effect is difficult to avoid when the objective distribution in the

696 JPD (Table 2) were constructed to lead to mainly high conditional probabilities.

697

698 Table 6

699 *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) |

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

703

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

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

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

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

733

734 Table 7

735 *The Probabilities of Drawing Each Inference in Experiments 2 and 3*

736

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

737 *Note: Inf. = Inference; Manip. = Manipulation. The same probability distribution was used in*
738 *Experiment 3, where it implements the inferences and manipulations shown in parentheses.*
739

740 **Method**

741 **Participants.** 334 participants were recruited via **MTurk** after some were excluded
742 because they may have been duplicates or participated in Experiment 1. All participants who
743 completed the experiment received a small payment (between US\$0.50 and US\$1.00). 53.6%
744 were female and the sample was aged between 18 and 83 with a median age of 36 years (mean =
745 39.44, SD = 13.32). English was the first language of 96.4% of participants.

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

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

762

763 **Results and Discussion**

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

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

779

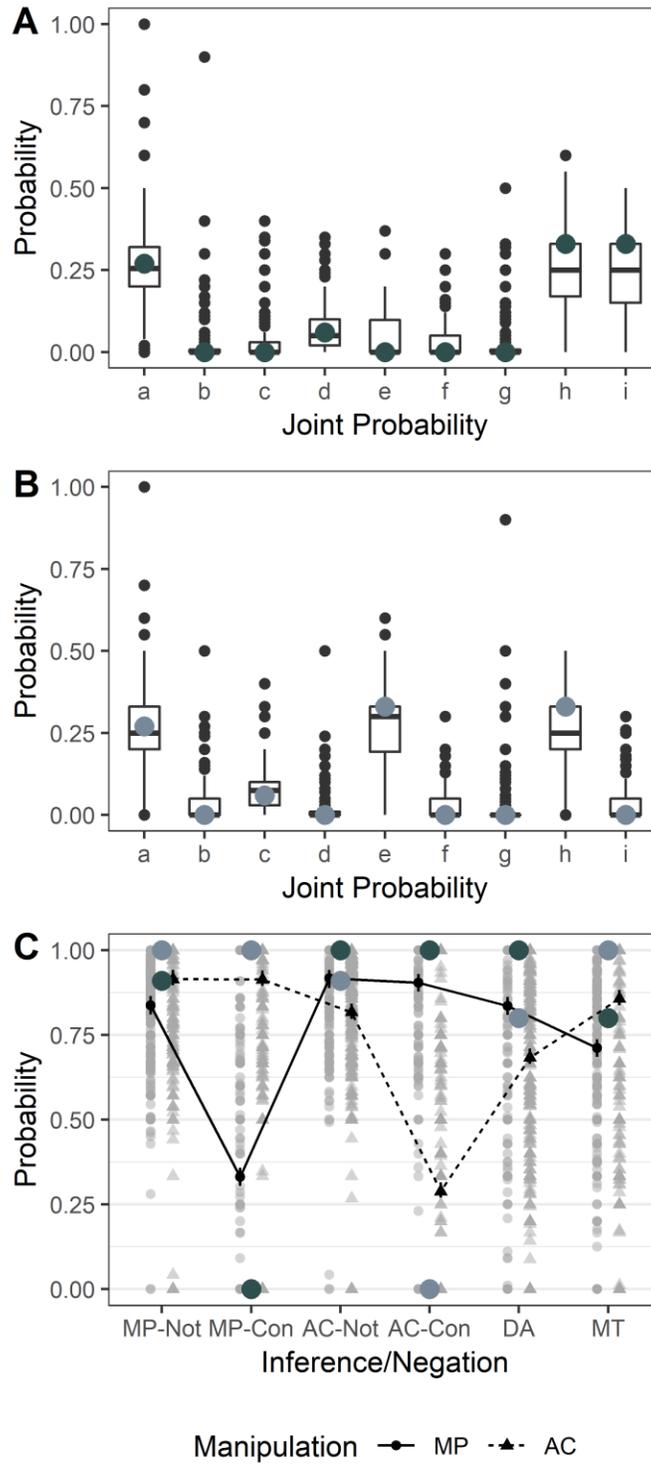
780

781

782

783

784 Figure 3
 785 *Joint Probabilities and Calculated Conditional Probabilities from the Probability Verification*
 786 *Task in Experiment 2*



787

788 *Note. A. Box-plots for the verification judgements for all cells of Table 6:MP-Manipulation. B.*
 789 *Box-plots for the verification judgements for all cells of Table 6:AC-Manipulation. C. Mean*
 790 *calculated conditional probabilities for each inference based on the estimates shown in panels A*
 791 *and B, error bars = 95% HDI. In these panels, the large dark grey points indicate the objective*
 792 *probabilities for the MP-Manipulation and the large light grey points indicate the objective*
 793 *probabilities for the AC-Manipulation.*
 794

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

802 For the MP-manipulation, MP-Con (mean = .33 [.28, .37]) was lower than MP-Not
 803 (mean = .84 [.79, .88]), $\bar{d} = 18.67 [16.80, 20.69]$, $1.0 \notin ROPE$), but zero was a credible value for
 804 the difference between AC-Con (mean = .90 [.87, .94]) and AC-Not (mean = .92 [.88, .92]), \bar{d}
 805 = .63 [-1.12, 2.32], $.70 \notin ROPE$). These results reversed for the AC-manipulation, zero was a
 806 credible value for the difference between MP-Con (mean = .91 [.86, .97]) and MP-Not (mean
 807 = .91 [.86, .97]), $\bar{d} = .03 [-1.81, 1.90]$, $.46 \notin ROPE$), but AC-Con (mean = .29 [.25, .33]) was
 808 lower than AC-Not (mean = .82 [.77, .87]), $\bar{d} = 19.01 [17.28, 20.59]$, $1.0 \notin ROPE$). We did not
 809 further analyze the results for DA and MT, but note that the calculated conditional probabilities
 810 followed the cross over pattern predicted by the objective values. In summary, the calculated
 811 conditional probabilities based on the verification task produced the predicted MP-manipulation
 812 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) \approx $\Pr(\neg p_1 | q_3)$

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

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

823 We also assessed the critical simple effects. For the MP-manipulation, the probability of
 824 endorsing MP-Con (mean = .68 [.60, .76]) was lower than MP-Not (mean = .97 [.96, .99]), $\bar{d} =$
 825 7.63 [5.60, 9.57], $1.0 \notin$ ROPE), but zero was a credible value for the difference between AC-Con
 826 (mean = .96 [.94, .98]) and AC-Not (mean = .94 [.91, .97]), $\bar{d} = -1.23 [-3.24, 1.00]$, $.81 \notin$
 827 ROPE). These results reversed for the AC-manipulation, zero was a credible value for the
 828 difference between MP-Con (mean = .94 [.92, .96]) and MP-Not (mean = .94 [.91, .96]), $\bar{d} =$
 829 $-.43 [-2.58, 1.76]$, $.58 \notin$ ROPE), but AC-Con (mean = .55 [.50, .60]) was lower than AC-Not
 830 (mean = .93 [.91, .96]), $\bar{d} = 15.37 [13.23, 17.61]$, $1.0 \notin$ ROPE).

831

832

833

834

835

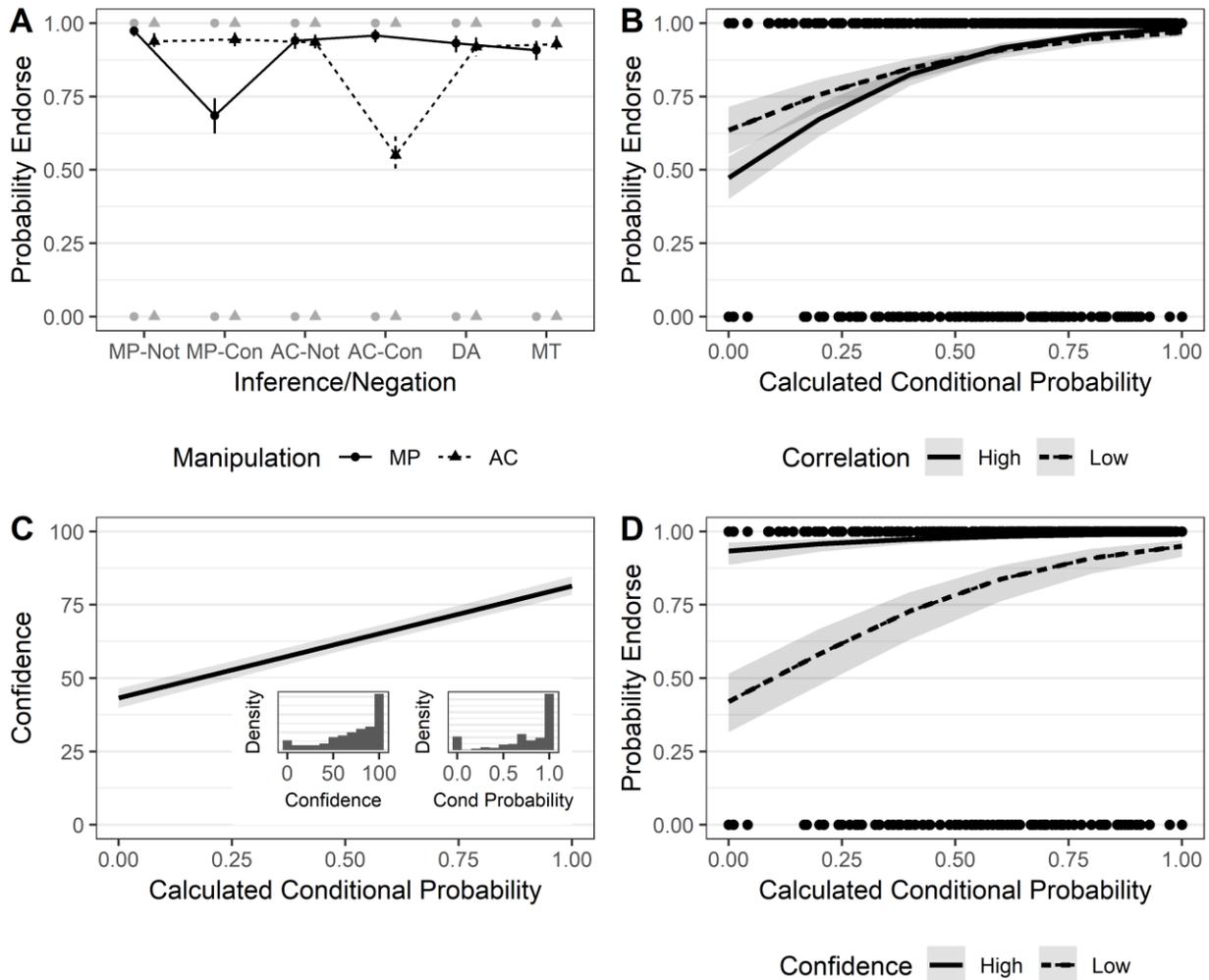
836

837

838

839

840 Figure 4
 841 *The Results of the Inference Tasks in Experiment 2*



842

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

850

851 In this experiment, we observed the predicted interactions confirming Hypothesis 1. An
 852 implicit negation effect only occurs when the contrast set member used to implicitly negate the
 853 antecedent or consequent indicates a low conditional probability of the conclusion. This analysis

854 directly addresses the possible criticism of Experiment 1 that we observed these effects only for
 855 the high correlation group. In analyzing these key predictions, in this experiment and the next,
 856 we did not split participants by high or low correlation groups.

857

858 Table 8

859 *Model Comparison for Predicting Inference Endorsement Rates in Experiment 2*

| | <i>LOOIC</i> | <i>SE</i> | <i>k</i> | Δ <i>LOOIC</i> | Δ <i>elpd</i> | Δ <i>se</i> | <i>Weight</i> |
|------------------------|--------------|-----------|----------|-----------------------|----------------------|--------------------|---------------|
| 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 |

860

861 *Notes. M1: Endorse* \sim *Inf*Neg* + (*Inf*Neg*|*PaGr*), *M2: Endorse* \sim *Inf* + *Neg* + (*Inf* + *Neg*|*PaGr*).
 862 *Estimated number of parameters (k), the difference (Δ LOOIC), the difference in expected log*
 863 *posterior predictive density (Δ elpd) and its standard error (Δ se), and the Bayesian stacking*
 864 *weights (LOOIC-weight).*

865

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

871

872 Table 9.

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

| | <i>LOOIC</i> | <i>SE</i> | <i>k</i> | Δ <i>LOOIC</i> | Δ <i>elpd</i> | Δ <i>se</i> | <i>Weight</i> |
|----|--------------|-----------|----------|-----------------------|----------------------|--------------------|---------------|
| 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 |

875

876 *Notes. M3: Endorse ~ Cond*Corr + (I|Participant) + (Cond*Corr|PaGr), M4: Endorse ~ Corr*
 877 *+ (I|Participant) + (Corr|PaGr), M5: Endorse ~ InfNeg*Manip + (InfNeg*Manip|PaGr).*
 878 *Estimated number of parameters (k), the difference in LOOICs (Δ LOOIC), the difference in*
 879 *expected log posterior predictive density (Δ elpd) and its standard error (Δ se), and the Bayesian*
 880 *stacking weights (LOOIC-weight).*

881

882 Table 9 shows the results of the model comparison. The stacking weights and Δ elpd
 883 converged on identifying M3 as the best model, confirming the results of Experiment 1 that most
 884 information relevant to drawing these inferences is in the predicted conditional probabilities.
 885 Figure 4B shows the relation between calculated conditional probability and endorsement rates
 886 for the high and low correlation groups for M3. The slope for the high correlation group was
 887 65.57 [34.88, 100.81] ($b > 0$, $1.0 \notin$ ROPE), that is, a .1 increase in calculated conditional
 888 probability increases the odds that an inference will be endorsed by 6.60. For the low correlation
 889 group, the slope was 18.56 [8.88, 30.02] ($b > 0$, $1.0 \notin$ ROPE), that is, a .1 increase in calculated
 890 conditional probability increases the odds by 1.9. The intercept for the high correlation group
 891 was .92 [.59, 1.26], indicating that when the calculated conditional probability was zero, an
 892 inference was marginally more likely to be rejected than endorsed. For the low correlation group
 893 the intercept was 2.02 [1.10, 3.16]. The intercept was higher for the low correlation group than

894 for the high ($\bar{d} = -2.67 [-6.01, .01], .97 \notin \text{ROPE}$), and the slope was steeper for the high
895 correlation group than for the low ($\bar{d} = 3.57 [1.27, 6.30], 1.0 \notin \text{ROPE}$).

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

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

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

931

932 **Experiment 3: MT and DA Manipulation**

933 We have demonstrated that we can produce or eliminate the implicit negation effect by varying
934 the learned probabilistic structure of the relevant contrast sets for MP and AC. In Experiment 3,
935 we attempted to replicate and generalize these findings to the MT and DA inferences. In this
936 experiment, we also used abstract material to show that we can produce the same probabilistic
937 effects for the materials that first demonstrated the implicit negations effect. We used abstract
938 content involving shapes and colours. The same probability manipulation as in Table 6 achieves

939 the desired result using the conditional *if it is white, then it is a van*. The AC-manipulation then
940 generates an MT-manipulation and the MP-manipulation generates a DA-manipulation. We show
941 the probability of drawing each inference in Table 7. In Experiment 3, $p_1 = \text{red/white}$, $p_2 =$
942 yellow/blue , $p_3 = \text{blue/red}$, $q_1 = \text{circle/van}$, $q_2 = \text{triangle/car}$, and $q_3 = \text{square/motorbike}$.

943

944 **Method**

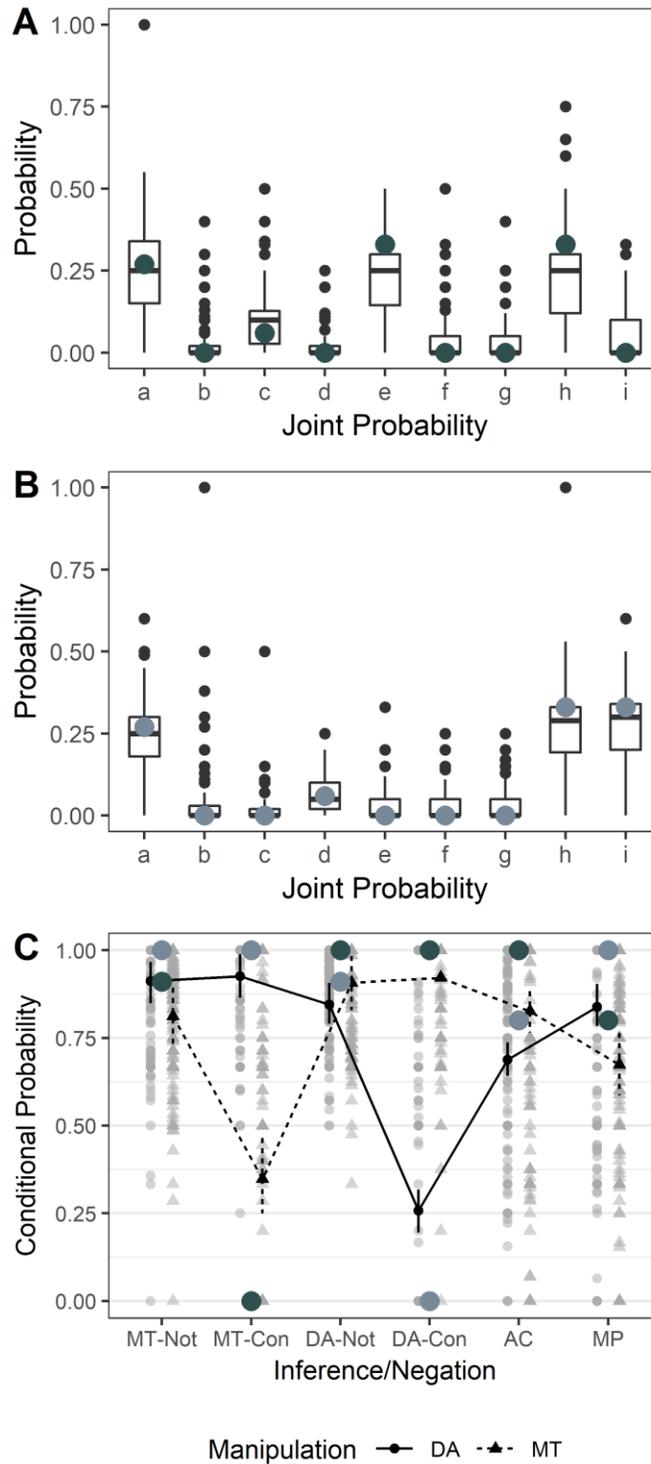
945 **Participants.** 168 participants were recruited via **MTurk** after some were excluded
946 because they may have been duplicates or participated in Experiments 1 or 2. All participants
947 who completed the experiment received a small payment (between US\$0.50 and US\$1.00).
948 56.0% were female and the sample was aged between 19 and 75 with a median age of 34 years
949 (mean = 38.05, SD = 13.75). English was the first language of 96.4% of participants.

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

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

961

962 Figure 5
 963 *Joint Probabilities and Calculated Conditional Probabilities from the Probability Verification*
 964 *Task in Experiment 3*
 965



967 *Note. A. Box-plots for the verification judgements for all cells of Table 6:MT-Manipulation. B.*
968 *Box-plots for the verification judgements for all cells of Table 6:DA-Manipulation. C. Mean*
969 *calculated conditional probabilities for each inference based on the estimates shown in panels A*
970 *and B, error bars = 95% HDI. In these panels, the large dark grey points indicate the objective*
971 *probabilities for the MT-Manipulation and the large light grey points indicate the objective*
972 *probabilities for the DA-Manipulation.*
973

974 **Results and Discussion**

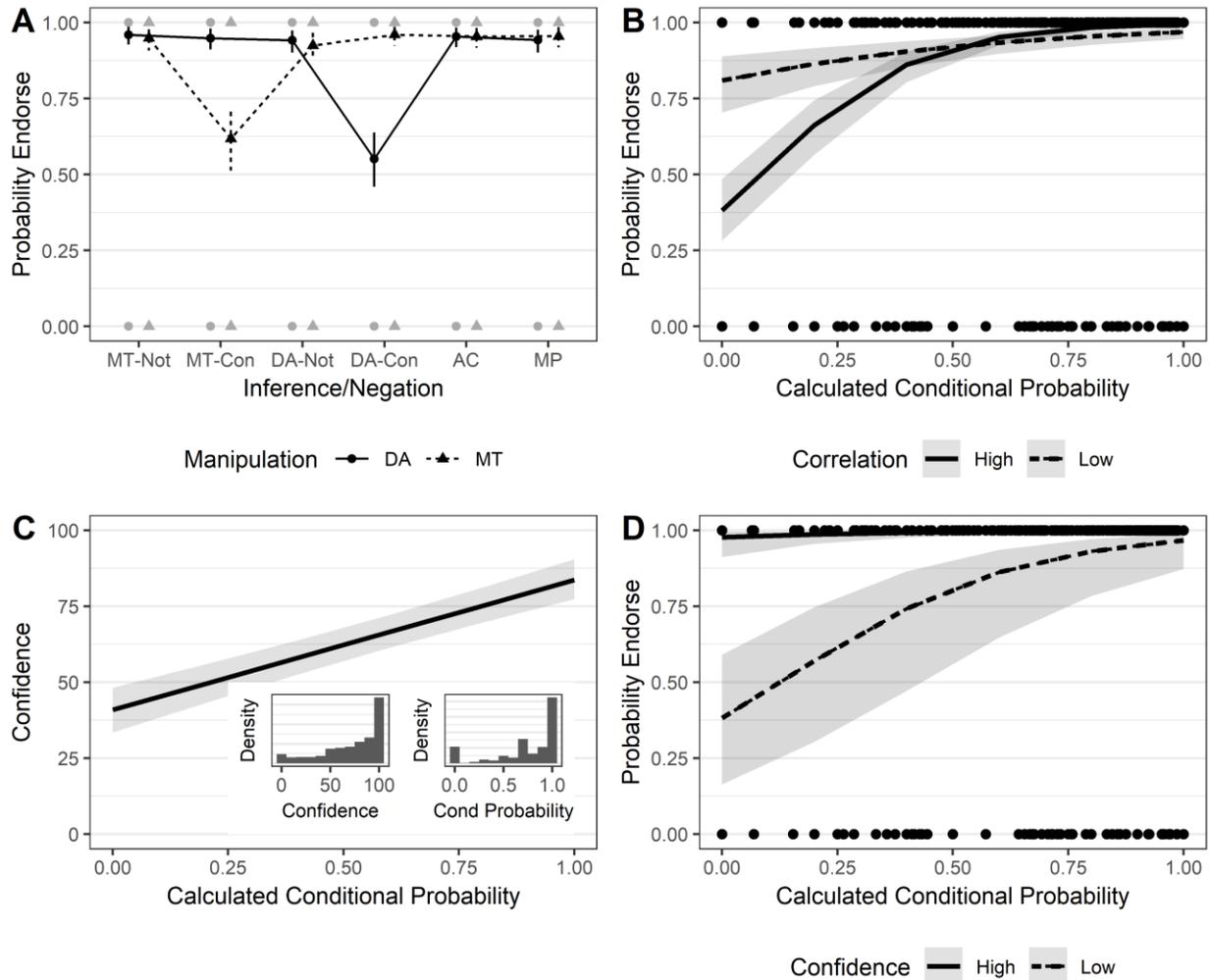
975 **Attention test.** The mean error rate (out of 120) was less than 1.0 % (1.10, SD = 4.24).
976 Most participants paid attention to the stimuli in the learning task and so we did not exclude any
977 participants from the subsequent analyses.

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

986 We made the same correction for missing values because of division by zero when
987 calculating conditional probabilities as in Experiments 1 and 2, which affected 19 participants
988 (11.3%) and 2.4% of cell values in participants subjective JPDs. Again, this correction did not
989 alter the correlations with the objective values. Figure 5C shows the estimated marginal means of
990 the calculated conditional probabilities for each inference split by manipulation (*Manip*). We
991 estimated these means using the same linear mixed model as in Experiment 2.

992

993 Figure 6
 994 *The Results of the Inference Tasks in Experiment 3*



995
 996 Notes. A. The probability of endorsing each inference for the MT- and DA-manipulations
 997 ($Endorse \sim InfNeg * Manip + (InfNeg * Manip | PaGr)$), error bars = 95% HDI; B. The probability
 998 of endorsing an inference predicted by the calculated conditional probability for the high and
 999 low correlation groups; C. The relationship between calculated conditional probability and
 1000 confidence for the high correlation group showing density plots for each variable; D. The
 1001 probability of endorsing an inference predicted by the calculated conditional probability for the
 1002 high correlation group with high and low confidence.

1003

1004 For the MT-manipulation, MT-Con (mean = .35 [.25, .46]) was lower than MT-Not
 1005 (mean = .81 [.73, .89]), $\bar{d} = 9.62 [7.12, 11.99]$, $1.0 \notin ROPE$), but zero was a credible value for
 1006 the difference between DA-Con (mean = .92 [.84, 1.00]) and DA-Not (mean = .91 [.83, .98]), \bar{d}

1007 = -.34 [-2.62, 1.92], .59 \notin ROPE). These results reversed for the DA-manipulation, zero was a
 1008 credible value for the difference between MT-Con (mean = .93 [.86, .99]) and MT-Not (mean
 1009 = .91 [.85, .97]), \bar{d} = -.45 [-2.84, 1.80], .62 \notin ROPE), but DA-Con (mean = .26 [.20, .32]) was
 1010 lower than DA-Not (mean = .84 [.79, .91]), \bar{d} = 19.47 [17.20, 22.14], 1.0 \notin ROPE). We did not
 1011 further analyze the results for AC and MP, but note that the calculated conditional probabilities
 1012 followed the cross over pattern predicted by the objective values. In summary, the calculated
 1013 conditional probabilities based on the verification task produced the predicted MT-manipulation
 1014 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) \approx $\Pr(\neg p_1 | q_3)$
 1015 (DA-Con) and the predicted DA-manipulation such that $\Pr(\neg q_1 | \neg p_1)$ (MT-Not) \approx $\Pr(\neg q_1 | p_2)$ (MT-
 1016 Con), and $\Pr(\neg p_1 | \neg q_1)$ (DA-Not) > $\Pr(\neg p_1 | q_3)$ (DA-Con).

1017 **Inference Tasks.** We observed no differences for the abstract materials and so we first
 1018 fitted the same model to the inference task as in Experiment 2 (see, Figure 6A: Notes for the
 1019 model) with the combined Path and Group variable as a random factor. We show the estimated
 1020 marginal means in Figure 6A. We then looked at the interaction between inference (*Inf*: MT and
 1021 DA) and negation (*Neg*: Not, Con) for each manipulation. As in Experiments 1 and 2, we
 1022 compared a model which included the interaction (M1) with one with only the main effects (M2)
 1023 (see, Table 10: Notes), and we show the results in Table 10. The stacking weights and Δelpd
 1024 converged on identifying M1, which includes the interaction, as the best model for both
 1025 manipulations.

1026 We also assessed the critical simple effects. For the MT-manipulation, MT-Con (mean
 1027 = .62 [.51, .71]) was lower than MT-Not (mean = .95 [.91, .98]), \bar{d} = 8.96 [6.34, 11.57], 1.0 \notin
 1028 ROPE), but zero was a credible value for the difference between DA-Con (mean = .96 [.92, .99])
 1029 and DA-Not (mean = .92 [.88, .97]), \bar{d} = -1.73 [-4.61, .87], .88 \notin ROPE). These results reversed

1030 for the DA-manipulation, zero was a credible value for the difference between MT-Con (mean
 1031 = .95 [.91, .98]) and MT-Not (mean = .96 [.93, .99]), $\bar{d} = .71 [-2.05, 3.51]$, $.66 \notin \text{ROPE}$), but
 1032 DA-Con (mean = .55 [.50, .60]) was lower than DA-Not (mean = .93 [.91, .96]), $\bar{d} = 11.10$
 1033 [8.34, 13.70], $1.0 \notin \text{ROPE}$).

1034

1035 Table 10

1036 *Model Comparison for Predicting Inference Endorsement Rates in Experiment 3*

| | <i>LOOIC</i> | <i>SE</i> | <i>k</i> | ΔLOOIC | Δelpd | Δse | <i>Weight</i> |
|------------------------|--------------|-----------|----------|----------------------|---------------------|-------------------|---------------|
| 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 |

1037

1038 *Notes. M1: Endorse ~ Inf*Neg + (Inf*Neg|PaGr), M2: Endorse ~ Inf + Neg + (Inf + Neg|PaGr).*
 1039 *Estimated number of parameters (k), the difference (ΔLOOIC), the difference in expected log*
 1040 *posterior predictive density (Δelpd) and its standard error (Δse), and the Bayesian stacking*
 1041 *weights (LOOIC-weight).*

1042

1043 Replicating Experiment 2, but now for MT and DA, we observed the predicted
 1044 interactions confirming Hypothesis 1. An implicit negation effect only occurs when the contrast
 1045 set member used to implicitly negate the antecedent or consequent indicates a low conditional
 1046 probability of the conclusion.

1047 **Calculated conditional probabilities.** We next tested whether the calculated conditional
 1048 probabilities (*Cond*) were good predictors of responses in the inference task (*Endorse*). We
 1049 compared the same models as in Experiment 2 (see Table 11: Notes for the models compared).
 1050 M5 is the model used to generate Figure 6A. Table 11 shows the results of the model
 1051 comparison. The stacking weights and $\Delta elpd$ converged on identifying M3 as the best model,
 1052 confirming the results of Experiments 1 and 2 that most information relevant to drawing these
 1053 inferences is in the predicted conditional probabilities. Figure 6B shows the relation between
 1054 calculated conditional probability and endorsement rates for the high and low correlation groups
 1055 for M3. The slope for the high correlation group was 365.68 [101.45, 716.17] ($b > 0$, $1.0 \notin$
 1056 ROPE), that is, a .1 increase in calculated conditional probability increases the odds that an
 1057 inference will be endorsed by 36.5. For the low correlation group, the slope was 8.09 [2.25,
 1058 15.30] ($b > 0$, $1.0 \notin$ ROPE), that is, a .1 increase in calculated conditional probability increases
 1059 the odds by .81. The intercept for the high correlation group was .64 [.33, 1.00], indicating that
 1060 when the calculated conditional probability was zero, an inference was marginally more likely to
 1061 be rejected than endorsed. For the low correlation group the intercept was 7.63 [2.63, 14.48]. The
 1062 intercept was higher for the low correlation group than for the high ($\bar{d} = -2.92 [-5.86, -.76]$, $1.0 \notin$
 1063 ROPE), and the slope was steeper for the high correlation group than for the low ($\bar{d} = 2.64 [.67,$
 1064 5.21], $1.0 \notin$ ROPE).

1065 Replicating Experiments 1 and 2, calculated conditional probability was the best
 1066 predictor of inference endorsement. This experiment also confirmed that correlation had a
 1067 moderating effect. With the stronger probability manipulation, better understanding of the
 1068 probability distribution (high correlation) leads to greater sensitivity (lower intercept, steeper

1069 slope). Replicating Experiment 2, the stronger probability manipulation led to reduced
 1070 uncertainty at the lower end of the scale, revealing that the intercepts also differed.

1071

1072 Table 11.

1073 *Model Comparison for Predicting Endorsement Rates from Calculated Conditional Probabilities*
 1074 *in Experiment 3*

| | <i>LOOIC</i> | <i>SE</i> | <i>k</i> | Δ <i>LOOIC</i> | Δ <i>elpd</i> | Δ <i>se</i> | <i>Weight</i> |
|----|--------------|-----------|----------|-----------------------|----------------------|--------------------|---------------|
| 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 |

1075

1076 *Notes. M3: Endorse ~ Cond*Corr + (1|Participant) + (Cond*Corr|PaGr), M4: Endorse ~ Corr*
 1077 *+ (1|Participant) + (Corr|PaGr), M5: Endorse ~ InfNeg*Manip + (InfNeg*Manip|PaGr).*
 1078 *Estimated number of parameters (k), the difference in LOOICs (Δ LOOIC), the difference in*
 1079 *expected log posterior predictive density (Δ elpd) and its standard error (Δ se), and the Bayesian*
 1080 *stacking weights (LOOIC-weight).*

1081

1082 **Confidence.** We next assessed the relationship between confidence and the predicted
 1083 conditional probabilities. Figure 6C shows that they are linearly related, which we again assessed
 1084 with separate intercepts for each participant and *PaGr* as a random effect. The population slope
 1085 was 42.88 [30.80, 55.51] ($b > 0$, $1.0 \notin$ ROPE), indicating that a 0.1 increase in conditional
 1086 probability led to a 4.28 point rise in confidence. Both distributions were skewed to the high end
 1087 of the scale (see subplots in Figure 6C), and their median values were .88 (conditional
 1088 probability) and 83 (confidence). Figure 6D shows that, replicating Experiment 2, confidence did
 1089 not moderate the effect of conditional probability on inference endorsement. As for Experiment

1090 2, Figure 6D is explained by the high correlation between confidence and calculated conditional
1091 probability (Figure C6).

1092 **Summary.** Experiment 3 confirmed Hypothesis 1 for MT and DA. There was an implicit
1093 negation effect for MT but not for DA for the MT manipulation, and an implicit negation effect
1094 for DA but not for MT for the DA manipulation. Not only were the simple effects significant, a
1095 model containing the interaction was a more accurate predictor of the data than a model with
1096 only the main effects. The calculated conditional probabilities for each inference derived from
1097 participants' JPD estimates, were also the best predictor of the probability of endorsing an
1098 inference, confirming Hypothesis 3. Moreover, understanding the probability manipulation
1099 moderated the effect, with the high correlation group's inference endorsements showing greater
1100 sensitivity to calculated conditional probability (lower intercept, higher slope). In contrast,
1101 confidence, although highly correlated with calculated conditional probability, confirming
1102 Hypothesis 4, did not moderate its effect on inference endorsement. This result is consistent with
1103 previous research that treated judgements of confidence as proxies for probabilities. These results
1104 are not consistent with other theories, which all predict an implicit negation effect for both MT
1105 and DA regardless of the probability manipulation used in these experiments.

1106

1107 **General Discussion**

1108 Experiments 1 to 3 provided focused experimental tests of the new paradigm probabilistic
1109 explanation of the implicit negation effect in conditional inference. We used short discrete
1110 learning tasks to impart probabilistic information about contextually limited sets of objects and
1111 their properties to manipulate whether an implicitly negated premise would lead to a high or low
1112 conditional probability of the conclusion. In Experiment 1, for the high correlation group we

1113 observed an implicit negation effect for MP but not for AC, consistent with the probability
1114 manipulation. The effects were large in terms of effect size but not of the same apparent
1115 magnitude as previously observed. In Experiment 2, we strengthened the probability
1116 manipulation and added an AC manipulation to test whether we could elicit and suppress the
1117 effect for both inferences. This manipulation produced a much larger effect on calculated
1118 conditional probabilities and a correspondingly larger implicit negation effect. We also observed
1119 the key interaction showing an implicit negation effect only when predicted by the probability
1120 manipulation. Experiment 3 replicated these findings for MT and DA inferences. Across all three
1121 experiments, the calculated conditional probability was the best predictor of the odds of
1122 endorsing an inference and this effect was moderated by the strength of the correlation between
1123 people's judgements of the joint probabilities (Tables 2 and 6) and the objective values.
1124 Participants who had better learned the probability distribution (high correlation group) showed
1125 greater sensitivity (lower intercept, higher slope) to the calculated conditional probability when
1126 endorsing inferences. Calculated conditional probability predicted confidence in whether
1127 participants endorsed an inference or not, but confidence did not moderate its effect on inference
1128 endorsement. This result is consistent with previous research that used confidence judgements as
1129 proxies for probabilities. These results raise a number of issues that we now address. We begin
1130 by looking at Bayesian New Paradigm approaches that can implement the predictions that we
1131 have just tested.

1132

1133 **New Paradigm Probabilistic Approaches**

1134 In deriving our predictions we have assumed that the probability of the conclusion of an
1135 inference is the conditional probability of the conclusion given the categorical premise.

1136 However, as we indicated in the introduction, this rubric does not provide an account of what
 1137 people are doing when they learn the categorical premise that provides a theory of inference at
 1138 either the computational or algorithm level. Fortunately, as we also observed, both approaches
 1139 we now consider lead to exactly the same predictions that our experiments have just tested.

1140 **Belief revision.** One approach is to treat inference as belief revision by conditionalization
 1141 (Eva & Hartmann, 2018; Oaksford & Chater, 2007, 2010b, 2013). This approach provides a
 1142 computational level theory that justifies our predictions. As we have argued, learning from
 1143 experience or a reliable informant leads people to revise their degrees of belief from a
 1144 distribution like Pr_0 to new a distribution like Pr_1 in Table 1. Conditionalization similarly treats
 1145 learning the categorical premise as belief revision to a new distribution Pr_2 . By Jeffrey
 1146 conditionalization this is achieved via the law of total probability. For example, (2) shows how to
 1147 calculate the new probability of the conclusion for the MP inference, where you learn a new
 1148 probability of p , $Pr_2(p)$, that is you come to believe that *Johnny travelled to Manchester* more
 1149 strongly ($> .4$).

$$1150 \quad Pr_2(q) = Pr_1(q|p)Pr_2(p) + Pr_1(q|\neg p)Pr_2(\neg p) \quad (2)$$

1151 If, however, learning p leads to $Pr_2(p) = 1$ (perhaps you think your informant is completely
 1152 reliable, i.e., Johnny is definitely travelling to Manchester), then (2) reduces to Bayesian
 1153 conditionalization, where $Pr_2(\neg p) = 0$. Consequently, MP on the conditional *if p then q* in Pr_1 in
 1154 Table 1 leads to:

$$1155 \quad Pr_2(q) = Pr_1(q|p)Pr_2(p) = Pr_1(q|p) = .75 \quad (3)$$

1156 That is, the new probability of the conclusion is the old conditional probability of the conclusion
 1157 given the categorical premise. Consequently, treating inference as Bayesian conditionalization
 1158 justifies all our predictions.

1159 However, it could be argued that there is a problem with this approach. Take MT on Pr_1
 1160 in Table 1, which leads to (4).

$$1161 \quad Pr_2(\neg p) = Pr_1(\neg p|\neg q)Pr_2(\neg q) = Pr_1(\neg p|\neg q) = .833 \quad (4)$$

1162 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
 1163 find the conditional premise acceptable. That the probability of the conditional premise is not
 1164 invariant across the belief update means that it is difficult to regard the revision to Pr_2 as
 1165 capturing what it means to draw these inferences. This set of four logical inferences concern
 1166 what follows from the premises assumed true or highly probable. Indeed, given (4), this
 1167 approach seems to imply that we should now believe that Johnny never travels anywhere by
 1168 train.

1169 However, this argument turns on an equivocation between our enduring beliefs versus
 1170 how they allow us to draw inferences from the momentary and changing flow of information we
 1171 experience. Learning about the conditional premise involves adjusting your enduring beliefs
 1172 about Johnny's travelling habits (the transition from Pr_0 to Pr_1). However, learning the
 1173 categorical premise in inference does not have this effect. In this example, Pr_1 represents your
 1174 enduring beliefs about Johnny's travelling habits, however acquired. In contrast, Pr_2 concerns
 1175 how you revise your beliefs about a specific journey based on this knowledge, in which you
 1176 learn he travelled to Manchester, or he did not take the train, and so on. So what remains
 1177 invariant in the revision from Pr_1 to Pr_2 is the target conditional probability, $Pr(\neg q|\neg p)$ for
 1178 DA...etc. However, this revision, required for inference, does not mean that people abandon
 1179 their enduring beliefs about Johnny's travelling habits in Pr_1 . Although nothing intrinsic to
 1180 probability theory enforces this distinction, it is enforced in algorithms for implementing
 1181 probabilistic inference, for example, Bayes nets.

1182 **Bayes nets.** A simple Bayes net implementing the JPD Pr_1 in Table 1, consists of two
 1183 nodes, p and q , corresponding to Bayesian random variables each with two possible states, 1
 1184 (True) and 0 (False), and a directional link from p to q . Inference over the net consists of variable
 1185 instantiation, that is, setting p or q to one of their states, say, $p = 1$, and belief propagation across
 1186 the link to the q node or backwards to the p node. The probability that the q node is in either of
 1187 its two states is determined by its conditional probability table (CPT), which includes $Pr(q = 1|p$
 1188 $= 1) = .75$ (and so $Pr(q = 0|p = 1) = .25$) and $Pr(q = 1|p = 0) = .167$ (and so $Pr(q = 0|p = 0)$
 1189 $= .833$). Together with the marginal for p , $Pr(p = 1) = .4$, the parameters $Pr(q = 1|p = 1) = .75$,
 1190 and $Pr(q = 1|p = 0) = .167$ implements the JPD Pr_1 in Table 1 in the network. These parameters
 1191 encode our enduring beliefs about Johnny's travelling habits and remain invariant across
 1192 different instantiations of its variables to their states.

1193 In this framework, the evidence provided by the categorical premise need not persuade us
 1194 that, for example, the probability that Johnny travels to Manchester is 1, $Pr(p) = 1$, and so we
 1195 should now believe he travels nowhere else. Rather it provides hard evidence to instantiate p to
 1196 1, and to read off the probability that $q = 1$, in an MP inference. Hard evidence always
 1197 instantiates a variable to just one of its states. This process is like performing a Ramsey test,
 1198 supposing the categorical premise by instantiating the relevant state of a random variable,
 1199 adjusting (i.e., forward and backward belief propagation), and then reading off the probability of
 1200 the conclusion, which for MP will be the conditional probability $Pr(q = 1|p = 1)$. This process is
 1201 the same for the remaining inferences by forward (MP, DA) or backward belief propagation (MT,
 1202 AC). Like Bayesian conditionalization, it also justifies all our predictions and can be extended to
 1203 provide an algorithmic level account of inference with contrast sets.

1204 **Bayes nets, negative evidence, and contrast sets.** We can implement the JPD in Table 2
 1205 in a Bayesian network with ternary, rather than binary states, with the CPT in Table 12. This CPT
 1206 contains two random variables p (travel destinations) and q (modes of transport) with states $\{p_1,$
 1207 $p_2, p_3\}$ and $\{q_1, q_2, q_3\}$ respectively. The assertion *Johnny did not travel to Manchester* ($p = \neg p_1$),
 1208 does not provide hard evidence concerning to which other destination, Paris or Dublin, he did
 1209 travel. Rather, it provides negative evidence that p can only be instantiated to states p_2 or p_3 but
 1210 not to p_1 (Bilmes, 2004; Mrad, Delcroix, Piechowiak, Leicester, Mohamed, 2015; Pearl, 1988).

1211
 1212 Table 12
 1213 *Conditional probability table for a Bayes Net with ternary states implementing the JPD in Table*
 1214 *2 showing the conditional probabilities $Pr(q_i|p_i)$ and marginals for p_i .*
 1215

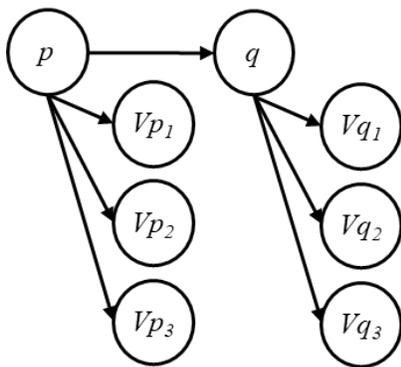
| | $p = p_1(.40)$ | $p = p_2(.16)$ | $p = p_3(.44)$ |
|-----------|----------------|----------------|----------------|
| $q = q_1$ | 0.750 | 0.625 | 0 |
| $q = q_2$ | 0.100 | 0.250 | 0.500 |
| $q = q_3$ | 0.150 | 0.125 | 0.500 |

1216
 1217 *Note:* $p_1 =$ Manchester, $p_2 =$ Paris, $p_3 =$ Dublin, $q_1 =$ train, $q_2 =$ ferry, $q_3 =$ plane.
 1218

1219 Following Pearl (1988), we can implement updating on negative evidence using virtual
 1220 nodes for each state of p and q . These virtual nodes are the children of the ternary nodes p and q
 1221 in a Bayes net (see Figure 7) with Table 12 as the CPT for the q node (see also, Bilmes, 2004;
 1222 Mrad et al., 2015). Figure 7 also shows the CPTs for the virtual nodes Vx_y . For the state p_1 of
 1223 node p $Pr(Vp_1 = 0|p = p_1) = 0$. Consequently, if $Vp_1 = 0$, then the travel destination (p) cannot be
 1224 Manchester (p_1), $p \neq p_1$. So the categorical premise *Johnny did not travel to Manchester* provides
 1225 evidence that $Vp_1 = 0$, and consequently that state p_1 is no longer a possible state of p but that
 1226 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

1227 net implements exactly the calculations we carried out over the JPD in Table 2 to derive our
 1228 predictions.¹⁰ Once this Bayes net is learned, inference is easy, and carried out by variable
 1229 instantiation and belief propagation, without the need for any conscious mental calculation. For
 1230 example, MP on (1), with the categorical premise *Johnny did not travel to Manchester*, involves
 1231 instantiating $Vp_1 = 0$, updating the network, and reading off the probability that $Vq_1 = 0$.¹¹

1232
 1233 Figure 7
 1234 Bayes Net implementing the CPT in Table 12 with virtual nodes implementing updating on
 1235 negative evidence
 1236



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 |

1237
 1238
 1239 It could be argued that this Bayes net would only work well for small contrast sets.
 1240 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.

1241 pragmatic factors will strongly constrain the contrast set, this may be all that is needed (Oaksford
 1242 & Stenning, 1992). Moreover, as we have argued (see introduction to *Experiment 2*), in inference
 1243 people only build very limited small-scale generative models related to their immediate deictic or
 1244 linguistic context (Oakford & Chater, 2020a).¹² These models are constructed on the fly (Chater,
 1245 2018) based on linguistic information and prior knowledge, in particular, from immediate past
 1246 experience, as in decision by sampling models (Stewart, et al., 2006).

1247 The Bayes net in Figure 7 also captures many of our intuitions about contrast sets; in
 1248 particular, that their internal probabilistic structure will render some contrast set members more
 1249 likely than others. Take the following examples with the word in bold stressed in speech.

1250 Johnny did not travel to **Manchester** by train (5)

1251 Johnny did not travel to Paris by **train** (5')

1252 The **cat** was not black (5'')

1253 The cat was not **black** (5''')

1254 In (5) Johnny travelled somewhere else by train, not Manchester, in (5'') Johnny travelled to
 1255 Paris by some other mode of transport, not train, in (5''') some other animal was black, not the
 1256 cat, and in (5''') the cat was some other colour, not black. Identifying the most likely contrast set
 1257 member for destination (5) involves instantiating p to $\neg p_1$, on negative evidence, and q to q_1 . The
 1258 model then identifies Paris as the most likely contrast set member, because $\Pr(p = p_2 | \neg p_1 = 0, q =$
 1259 $q_1) = 1$ and $\Pr(p = p_3 | \neg p_1 = 0, q = q_1) = 0$. In (5'), the model identifies ferry as the most likely
 1260 contrast set member because $\Pr(q = q_2 | p = p_2, \neg q_1 = 0) = .67$ but $\Pr(q = q_3 | p = p_2, \neg q_1 = 0) = .33$.
 1261 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.

1262 in Figure 7 may provide a more general theory of contrary negation and the effects of negative
1263 focus in speech.

1264 **Causal Bayes nets.** We have previously argued that people mentally represent
1265 conditionals in causal Bayes nets (Ali, Chater, & Oaksford, 2011; Ali, Schlottman, Shaw, Chater,
1266 & Oaksford, 2010; Chater & Oaksford, 2006; Oaksford & Chater, 2010b, 2013, 2016, 2017).
1267 However, to capture the implicit negation effect, we have not needed to assume any general
1268 probabilistic independencies and so the Bayes net in Figure 7 has been sufficient.¹³ However,
1269 our account of how people compute contrast sets borrows partly from causal approaches to
1270 category structure, in which intrinsic properties of a category cause the various features it
1271 possesses (Rehder, 2003a, 2003b, 2017). Moreover, we have suggested that people think about
1272 habits like causes, so, for Johnny, travelling to Manchester causes him to travel by train
1273 (Oaksford & Chater, 2010, 2020b). We may acquire habits and dispositions from our parents,
1274 peers, culture or by intention, but they are rapidly sedimented into the unconscious causes of our
1275 actions. All the elements of the ad hoc superordinate category (Barsalou, 1983)—places to which
1276 Johnny travels (p)—are causally related to travel destinations considered as features (q). It is a
1277 desiderata, therefore, to investigate models integrating CBNs with negative evidence in
1278 modelling conditional reasoning.

1279 A minor complication is that if we model contrast sets causally then the direction of
1280 causality matters. Some of our materials were diagnostic conditionals, for example, in the
1281 vehicles materials the conditional was *if it is not white, then it is not a van*. We think of objects
1282 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 Pr₁ in Table 1.

1283 causes its colour.¹⁴ A CBN representation would require representing the consequent (q) as the
 1284 cause and the antecedent (p) as the effect. This complication is minor, because we already know
 1285 from their patterns of discounting and augmentation inferences that people recode diagnostic
 1286 conditionals in this way (Ali et al., 2011).

1287 A possible argument against the appeal to CBNs, concern recent demonstrations that
 1288 people violate the independence assumptions of these models (Rehder, 2014; Rottman & Hastie,
 1289 2016). However, there are models that can account for these violations (Rehder, 2018).
 1290 Moreover, the empirically most adequate model may arise from limited sampling from initially
 1291 preferred states of the underlying generative causal model (Davis & Rehder, 2017; Rehder,
 1292 2018). It remains to be seen whether similar violations occur when identifying contrast set
 1293 members, but the theoretical machinery may be in place to explain them. Processing accounts
 1294 based on limited sampling from an underlying generative model have also been used to explain
 1295 away a variety of other biases (Dasgupta, et al., 2017; Hattori, 2016; Sanborn & Chater, 2016;
 1296 Stewart, et al., 2006)

1297

1298 **Alternative Theories**

1299 There are three alternative theories of the implicit negations effect, the matching heuristic
 1300 (Evans, 1998; Thompson, Evans & Campbell, 2013), mental models theory (MMT; Johnson-
 1301 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).

1302 contrast set hypothesis (Schroyens & Schaeken, 2000; Schroyens, Verschueren, Schaeken, &
 1303 d'Ydewalle, 2000). MMT implements the double hurdle theory proposed by proponents of the
 1304 heuristic approach. Consequently, these theories stand, and fall, together. The first hurdle is to
 1305 see an implicit negation as relevant, that is, as an instance of the negated antecedent or
 1306 consequent of a conditional.¹⁵ In MMT, negations are represented using explicit contradictory
 1307 negation tags. The first hurdle is that, unless people can recode the implicitly negated categorical
 1308 premise using such a tag, they do not realize that a constituent in a mental model has been denied
 1309 or affirmed. The second hurdle requires a double negation inference, so MT on (1), requires the
 1310 inference from *it is not the case that he did not travel to Manchester* ($\neg\neg p$) to *he travelled to*
 1311 *Manchester* ($\neg\neg p \rightarrow p$). This inference is only required for DA and MT. Both theories locate the
 1312 problem with implicit negations solely as a difficulty in seeing them as denying or affirming a
 1313 negated antecedent or consequent. Consequently, they do not predict any of the probabilistic
 1314 effects we observed.

1315 Binary sets, where there are, say, just two letters $\{A, K\}$ and the contrast set is a singleton,
 1316 remove the implicit negation effects in comparison to larger sets $\{A, K, W\}$ where the contrast set
 1317 has more than one member (Schroyens, Schaeken, Verschueren, & d'Ydewalle, 2000). The
 1318 *cardinality of the contrast set* hypothesis (CCS) is that a contrast set with more than one member
 1319 causes the implicit negation effect. According to this hypothesis with larger contrast sets,
 1320 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*.

1321 category, letters that are *not A*. Schroyens et al. (2000) observed implicit negation effects for
1322 contrast sets with two or more members (overall sets sizes of three or more) but not for
1323 singleton sets. Although CCS exploits the notion of a contrast set, it does not appeal to their role
1324 in computing probabilities. All the contrast sets in our experiments had two members.
1325 Consequently, our probabilistic manipulations removed the implicit negation effect even for
1326 contrast sets whose cardinalities were greater than one (we refer to this situation as “contrast
1327 set(s) > 1”), which is not consistent with the CCS hypothesis. We now briefly consider some
1328 recent further evidence supportive of the matching heuristic or mental models.

1329 In the Wason selection task, the matching heuristic response (see, *Footnote 15*) seems
1330 meta-cognitively fluent (Thompson, et al., 2013). That is, participants’ “answers consistent with
1331 a matching heuristic (i.e., selecting cards named in the rule) were made more quickly than other
1332 answers, were given higher FOR [feeling of rightness] ratings, and received less subsequent
1333 analysis as measured by rethinking time and the probability of changing answers” (Thompson, et
1334 al., 2013, p. 431). From a probabilistic perspective, this is not surprising as the probabilistic
1335 contrast set account makes the same predictions in this evidence acquisition task (Oaksford &
1336 Chater, 2003; 2007; Oaksford, Chater, Grainger & Larkin, 1997). It, therefore, provides a
1337 rational analysis of why in data acquisition a matching heuristic is rational. The question of
1338 whether this rational analysis is implemented by a heuristic or a probabilistic algorithm depends
1339 on whether behaviour can be changed by probabilistic manipulations and the results show that
1340 this is possible (e.g., Oaksford et al., 1997). We know of no similar demonstration of fluency for
1341 the matching responses in conditional inference. However, we would speculate that if people
1342 deploy such a heuristic in the conditional inference task, it is probably learned rather than hard-
1343 wired and so can be overridden by subsequent learning, as our experiments demonstrated.

1344 The motivation for an explicit negation tag in MMT derives from the psycholinguistic
1345 literature where it is hypothesized that people construct two representations of a negated
1346 assertion like “the door is not open” (Kaup, Zwaan, & Lüdtke, 2007; Khemlani et al., 2012,
1347 Orenes, Beltran, & Santamaria, 2014). In the first representation, the door is open and in the
1348 second, it is closed. This strategy works for binary opposites or antonyms, like open and closed,
1349 but what about “the dot is not blue” presented in an array of four coloured dots (Orenes et al.
1350 2014)? Here the second representation would have to include all the other three dots. The
1351 negations tag therefore acts as a short hand for the opposites when the overall set size is greater
1352 than two. If people represent opposites (contrast sets) for the contrast set > 1 case using a
1353 negations tag, then the content of both representations still includes the affirmative statement
1354 (e.g., blue dot). Using a visual world array like this, Orenes et al. (2014) used an innovative eye
1355 tracking experiment to show that visual attention switches to the alternative when sets are binary
1356 (singleton contrast set) but remains on the affirmative item when the contrast set > 1 . A finding
1357 that is consistent with the use of a negation tag for non-binary opposites.

1358 There are several points to make. First, in these visual world tasks, participants did not
1359 have to draw inferences, nothing depended on what the contrast set members might predict.
1360 Second, unlike our more real world materials, the contrast sets had no probabilistic structure. So,
1361 if the coloured dot was not blue it was equally likely to be one of the other three dots in the
1362 display. In our materials, for example, if Johnny did not travel to Manchester, he was far more
1363 likely to travel to Dublin than to Paris. Third, our experiments showed that people do not seem
1364 to have any trouble representing structured contrast sets with more than one member and
1365 drawing appropriate inferences over whatever mental representations of this situation they
1366 construct. Fourth, it also seems theoretically incongruous to argue that people automatically

1367 recode contrasts sets > 1 with negation tags but also argue that the use of a member of a contrast
1368 set > 1 to deny (affirm) a (negated) proposition causes a recoding problem. If people
1369 automatically recode these sets with negations tags, then why do they not automatically recode
1370 members of one of these sets when encountered in inference? If these contrast sets are
1371 *automatically* recoded with a negation tag, then the first hurdle in the mental model
1372 implementation of double hurdle theory has been jumped. Moreover, the second hurdle, double
1373 negation inferences for MT and DA, is probably a red herring. Our mini meta-analysis showed
1374 strong implicit negations effects also for MP and AC (see the introduction to *Experiment 1*),
1375 which our experiments replicated.

1376 Although it is unclear how it could integrate with the MMT account of the implicit
1377 negation effect, MMTs have been extended to capture probabilistic effects by annotating the
1378 possibilities they represent with probabilities (Johnson-Laird, Legrenzi, Girotto, Legrenzi, &
1379 Caverni, 1999). To model the current data this would involve representing the nine possible
1380 states in the JPDs in Tables 2 and 6 and their associated probabilities. The resulting mental model
1381 would be a notational variant of these tables. People would then have to calculate the relevant
1382 conditional probabilities by summing over the annotations to the relevant models (cells) and
1383 using the ratio formula (see *Footnote 11*). Prima facie, it seems unlikely that people are
1384 performing these calculations during inference, rather than compiling a representation as in
1385 Figure 7 during learning. Of course, because either theory would predict the same subjective
1386 calculated conditional probabilities they would predict the odds of people endorsing an inference
1387 equally well. The problem for MMT is that this is not its theory of the implicit negation effect.
1388 Moreover, it proposes an implausibly direct implementation of the joint probability distributions
1389 in Tables 2 and 6 and of the operations defined over them.

1390 We do not need to deny that our mental representations use negation tags on occasion. As
1391 we have pointed out, identifying contrast sets does not exhaust the way people used negations in
1392 natural language (Horn, 1989), and some may require people to represent information with a
1393 negation tag. We would argue, however, that our normally shallow knowledge of the world (Keil
1394 & Rozenblit, 2004; Sloman & Fernbach, 2017), like someone's knowledge of Johnny's
1395 travelling habits, means that most contrast sets are not large and are not much like the abstract
1396 domains of letters, numbers or coloured dots.

1397

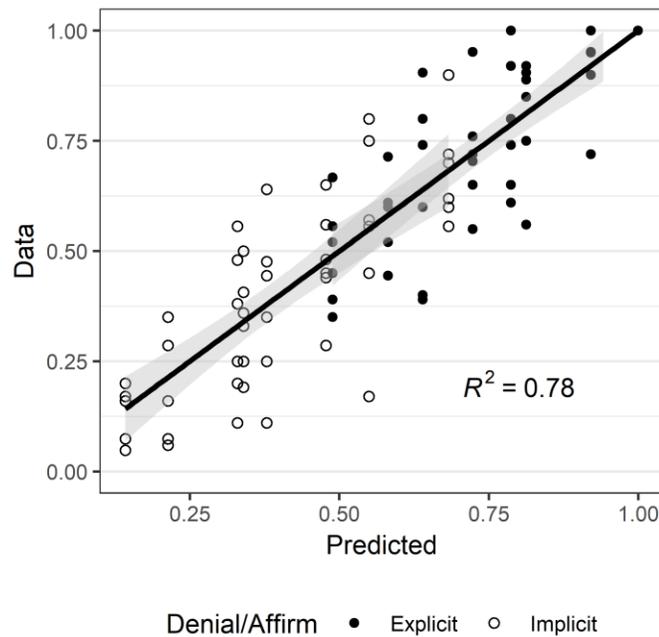
1398 **Modelling the Default Prior Pro.**

1399 Our focus has been on showing that targeted experimental manipulations of probabilities can
1400 produce or remove the implicit negation effect. However, can our account model the original
1401 implicit negations effect? The data have been reported in two different ways. Evans and Handley
1402 (1999) contrast whole tasks using explicit negations only (the explicit negations paradigm) with
1403 whole tasks using implicit negations only (the implicit negations paradigm). Eight of the possible
1404 sixteen conditions can reveal implicit negations effects. For example, MP on *if $\neg p_1$ then q_1* can
1405 use an explicit, $\neg p_1$, or an implicit, p_2 , categorical premise. The implicit paradigm alone also has
1406 eight conditions that reveal implicit negations effects (Schroyen et al. 2000). For example, MP
1407 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
1408 use a contrast set member p_2 to assert the negative antecedent. Both cases produce an implicit
1409 negations effect. For the same inference (e.g., MP) endorsements of the conclusion (q_1) fall
1410 compared to using the explicit negation ($\neg p_1$) on the same rule (*if $\neg p_1$ then q_1*) or the affirmative
1411 (p_1) on a different rule (*if p_1 then q_1*) where the target clause is affirmative. Here we modelled the
1412 data from the implicit negations paradigm.

1413 We modelled the six implicit negations paradigm conditions in Evans and Handley (1999:
 1414 Experiments 1: conditions: no-pictures, pictures, & Experiment 3) and Schroyens et al. (2000:
 1415 Experiment 1: conditions: set sizes 3, 5, and 9). There were 131 participants and 96 data points.
 1416 There is one complication. We had to model each of the four rules as if they involved different
 1417 content. First, this is always the case experimentally because the intention was to see what
 1418 follows from each rule independently. Second, if the same content is used, as it has been in
 1419 examples apparently questioning the probabilistic interpretation (Schoyens & Schaeken, 2003),
 1420 various conceptual absurdities result (Oaksford & Chater, 2003b). Third, the probability
 1421 conditional does not allow certain pairs of conditionals to be true (or to have high probability) at
 1422 the same time. The probability conditional respects the law of conditional excluded middle. In
 1423 standard binary logic *if p then q* and *if p then ¬q* are consistent. They can both be true if the
 1424 antecedent is false. In contrast, for the probability conditional, for which $\Pr(\text{if } p \text{ then } q) = \Pr(q|p)$,
 1425 these conditionals cannot be true together because if $\Pr(q|p) = 1$, then $\Pr(\neg q|p) = 0$.¹⁶ So, if these
 1426 conditionals shared the same content then they cannot both have a high probability. The same
 1427 argument applies to the pair *if ¬p then q* and *if ¬p then ¬q*. Finally, the four conditionals in the
 1428 negations paradigm are also related by necessity and sufficiency. So, if they share content, then *if*
 1429 *p then q* suggests that *p* is sufficient for *q* and *if ¬p then ¬q* suggests that *p* is necessary for *q*. If
 1430 *p* is necessary and sufficient for *q* then this should affect endorsements of DA and AC, which
 1431 would now be valid inferences. In summary, using the same content creates unwanted
 1432 dependencies between the four conditionals that we can rule out only by using different content
 1433 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.

1434 Figure 8
 1435 *Modelling the Implicit Negation Effect*



1436

1437 We fitted the model using the minimal contrast set structure of two members (overall set
 1438 size = three) for both antecedent and consequent as in Tables 2 and 6. We modelled each
 1439 conditional separately thereby assuming different content. The parameters were the nine joint
 1440 probabilities ($a - i$), which, because they must sum to one, meant there were eight free
 1441 parameters, to model 24 data points. Because the data constitute six replications of 16 data
 1442 points, the best a model can do is predict the mean across replications. With this number of free
 1443 parameters, this was indeed the outcome of the model fitting (see, Figure 8), the model
 1444 accounted for 78% of the variance in the data (coefficient to determination $R^2 = .78$).

1445 Figure 8 also separates out the data points for which a contrast set member (implicit)
 1446 affirms a negative or denies an affirmative (unfilled dots) and those where the negated
 1447 constituent (explicit) affirms a negative or denies an affirmative (filled dots). Figure 8 shows that
 1448 the implicit data and the predicted conditional probability were always lower than the explicit
 1449 cases. So, the explicit cases (*if p then q, if p then $\neg q$*) for MP, always had higher probabilities of

1450 the conclusion/proportion of endorsements than the implicit cases (*if $\neg p$ then q , if $\neg p$ then $\neg q$*).
1451 We show the best fitting parameter values in the Appendix, Table A1. They will allow us to
1452 calculate various quantities to see whether these results conform to recent proposals about
1453 conditional inference called “inferentialism.”

1454 In summary, our account of the implicit negation effect can account for the original
1455 effects observed using all four rules in the negations paradigm. The fundamental insight is that
1456 the use of a contrast set member raises the possibility that it does not predict the conclusion as
1457 strongly as the explicitly negated categorical premise of a conditional inference. In this sense, the
1458 cardinality of the contrast set account is correct in that any contrast set > 1 will raise this
1459 possibility (Schroyens, et al., 2000). However, the internal probabilistic structure of the ad hoc
1460 categories suggested by the assertion of the conditional causes the effect, not a difficulty in
1461 recognizing the contrast set member as an instance of the negated category.

1462

1463 **Probabilities**

1464 The calculated conditional probabilities predicted the odds of endorsing an inference well.
1465 However, even for those participants who understood the probability manipulation (high
1466 correlation) very low probabilities still frequently led people to endorse an inference. We could
1467 not expect people’s subjective probabilities to track the objective probability manipulation
1468 exactly. On the Bayesian view of probabilities, they are always relative to what somebody knows
1469 or believes, so the general form of a subjective probability statement is $\Pr(p|B)$, where B stands
1470 for an individual’s background beliefs. People know more about the domains of animals and
1471 vehicles and their colours than is given in the probability-learning task. Although the subjective
1472 estimates did follow the objective probabilities quite well.

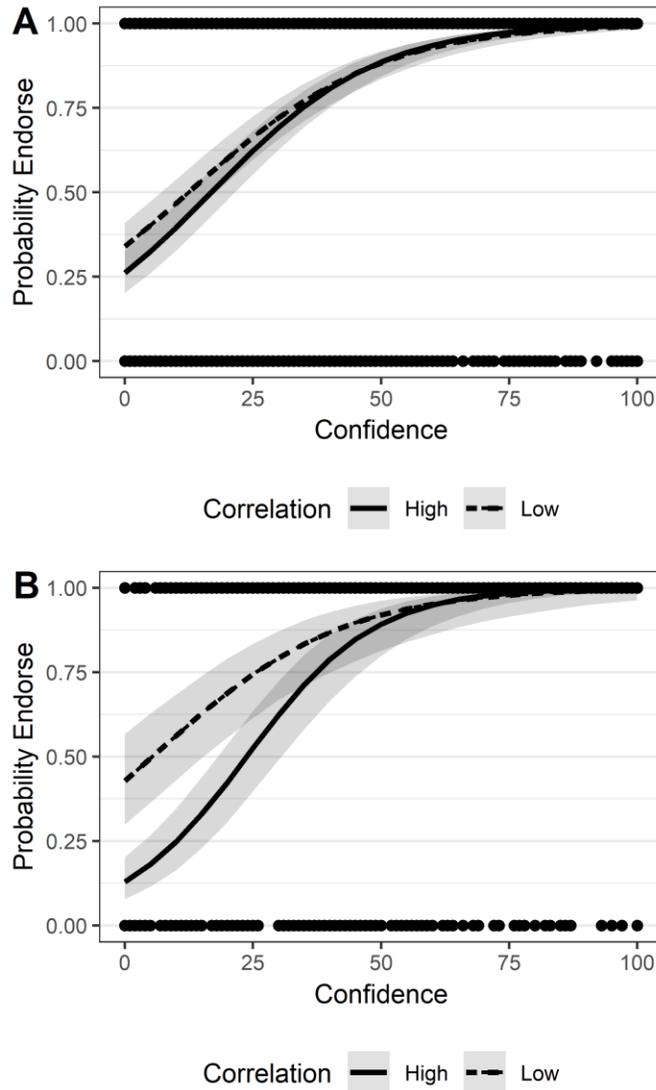
1473 One reason why endorsement rates may be high even for low calculated conditional
1474 probabilities, is that across all conditions the mean conditional probability was high at around 0.7
1475 (Expt. 1: Objective = .72, Subjective = .68(.19); Expt. 2: Objective = .71, Subjective = .75(.32);
1476 Expt. 3: Objective = .71, Subjective = .75(.32)). Consequently, on average, participants should
1477 endorse an inference, although this will depend on their personal criterion or cut-off. Moreover,
1478 they should endorse five out the six inferences they experienced in each manipulation, which
1479 again may bias participants towards endorsement. Given this potential bias toward endorsement,
1480 it is impressive that our results nonetheless showed a strong effect of calculated conditional
1481 probability on the odds of endorsing an inference.

1482 Another reason why the calculated conditional probabilities may not be better predictors
1483 of inference endorsement is the indirect method of computation and the reliance on the ratio
1484 formula to compute the conditional probabilities ($\Pr(q|p) = \Pr(p, q)/\Pr(p)$). The probability
1485 verification task is similar to versions of the probabilistic truth table task (Over et al, 2007). This
1486 task has been criticized as perhaps not revealing people's probabilistic interpretations of the
1487 conditional (Jubin & Barrouillet, 2019). The precise reasons do not matter, but an immediate
1488 response is that (a) these tasks (especially our task which involves filling in 9 cells of the JPD)
1489 creates a lot of room for error, and (b) the subjective Bayesian approach rejects the frequentist
1490 method and the ratio formula for calculating conditional probabilities. On the Bayesian
1491 interpretation, conditional probabilities are basic and suppositional, that is, they based on the
1492 Ramsey test (see, *Probabilities and Contrast Sets*).

1493
1494
1495
1496
1497
1498

1499 Figure 9

1500 *Predicting Endorsement Rates from Confidence for High and Low Correlation Groups*



1501

1502 *Notes: A: Experiment 2, B: Experiment 3. For both experiments the model fitted was $Endorse \sim$*
 1503 *$Conf * Corr + (1|Participant) + (Conf * Corr|PaGr)$.*

1504

1505 People’s probability judgements are more coherent when queried while drawing
 1506 inferences (Evans, Thompson, & Over, 2015). We have already shown that in our experiments,
 1507 calculated conditional probability directly predicts confidence in endorsing an inference.
 1508 Therefore, people’s confidence judgements, which we obtained when people are actually

1509 drawing inferences, may provide a more direct measure of the relevant conditional probabilities.
1510 As we have argued, during inference people effectively perform a Ramsey test, supposing the
1511 categorical premise to be true (see, *Bayes nets*). If their degree of belief in the conclusion goes
1512 above criterion, then they endorse the inference and report this degree of belief as how confident
1513 they are. If this is the right interpretation, then the suppositional account would predict that using
1514 confidence as a predictor should lead to a much steeper response curve showing sensitivity at
1515 both the high and the low ends of the scale. Moreover, if the probability-learning task has
1516 influenced people's subjective conditional probabilities as measured by the confidence
1517 judgements, then we would expect to see a moderating effect of high or low correlation (*Corr*).

1518 Figure 9 shows how the odds of endorsing an inference varied with confidence for the
1519 high and low correlation groups in Experiments 2 and 3. As predicted, the response curves are
1520 much steeper than for calculated conditional probability, and correlation in the probability
1521 verification task moderated the effect, especially in Experiment 3. Table 13 shows that in both
1522 Experiments 2 and 3, using confidence (M1) as a predictor yielded a much better fit to the data
1523 than calculated conditional probability (M2). However, even in the high correlation group in
1524 Experiment 3, people still seem biased to endorse an inference as revealed by the left-shift in the
1525 response curve (see, Figure 9). One would expect the odds of endorsing an inference to be one
1526 (probability = 0.5) when conditional probability was 0.5. As we observed, this may be because,
1527 on average, inferences in this task should be endorsed. De-biasing may be possible by balancing
1528 inferences so that equal numbers should be endorsed or rejected. The moderating effect of
1529 correlation demonstrates that the effects of the learning-phase endured to affect people's
1530 subjective probability judgements, as measured by confidence, in the inference tasks.

1531

1532

1533 Table 13

1534 *Model Comparison for Predicting Inference Endorsement Rates from Confidence vs. Calculated*
1535 *Conditional Probability*

| | <i>LOOIC</i> | <i>SE</i> | <i>k</i> | $\Delta LOOIC$ | $\Delta elpd$ | Δse | <i>Weight</i> |
|--------------|--------------|-----------|----------|----------------|---------------|-------------|---------------|
| 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 |

1536

1537 *Notes. M1: Confidence, M2: Calculated Conditional Probability. Estimated number of*
1538 *parameters (k), the difference ($\Delta LOOIC$), the difference in expected log posterior predictive*
1539 *density ($\Delta elpd$) and its standard error (Δse), and the Bayesian stacking weights (LOOIC-weight).*

1540

1541 **Inferentialism**

1542 A recent development in the psychology of reasoning is the realization that people tend to
1543 endorse conditionals only when they believe there is some kind of inferential link between the
1544 antecedent and the consequent. So for example, they do not regard conditionals like, *if the moon*
1545 *is made of cheese, Corbyn will be elected Prime Minister* as candidates for truth. Although, given
1546 that the moon is not made of cheese, we would logically have to endorse this conditional as true.
1547 This is one of the so-called “paradoxes of material implication.” There are two versions of
1548 inferentialism. On the semantic version, indicative conditionals express inferential or reason
1549 relations between the antecedent and consequent which are part of the truth conditions of the

1550 conditional (Douven, Elqayam, Singmann, & van Wijnbergen-Huitink, 2018; Douven &
 1551 Mirabile, 2018; Mirabile & Douven, in press). On the probabilistic version reason relations are
 1552 probabilistic and part of the acceptability conditions of indicative conditionals (Krzyżanowska,
 1553 Collins, & Hahn, 2017; Skovgaard-Olsen, Collins, Krzyżanowska, Hahn, & Klauer, 2019;
 1554 Skovgaard-Olsen, Kellen, Hahn, & Klauer, 2019; Skovgaard-Olsen, Kellen, Krahl, & Klauer,
 1555 2017; Skovgaard-Olsen, Singmann, & Klauer, 2016, 2017). Antecedent and consequent are
 1556 positively probabilistically relevant when $\Pr(q|p) > \Pr(q|\neg p)$, that is, when Delta-P (ΔP , Ward &
 1557 Jenkins, 1965) is positive. ΔP was found to moderate whether the Equation ($\Pr(\text{if } p \text{ then } q) =$
 1558 $\Pr(q|p)$) holds. Only when $\Delta P > 0$, that is, p and q are positively inferentially relevant, does the
 1559 Equation adequately predict whether a conditional is acceptable.

1560 The data from the probability verification task and the best fitting parameter values from
 1561 the model fits (see, *Modelling the default prior Pr_0*) allow us to check whether the materials in
 1562 these tasks show positive relevance. For Experiment 2, the objective probabilities for the *if* $\neg p$,
 1563 *then* $\neg q$ rule respected positive relevance. For the MP-manipulation, $\Delta P (\Pr(\neg q|\neg p) - \Pr(\neg q|p))$
 1564 = .91 , and for the AC-Manipulation, $\Delta P = .80$. Aggregating across manipulations, for the
 1565 subjective probabilities, mean $\Delta P = .64$ (SD = .36). Only 54 out of 668 calculated ΔP s (7.8%)
 1566 were zero or negative and 52 of these came from the low correlation group. For Experiment 3,
 1567 the objective probabilities for the *if* p *then* q rule respected positive relevance. For both the DA-
 1568 and the MT-manipulations, $\Delta P (\Pr(q|p) - \Pr(q|\neg p)) = .80$. Aggregating across manipulations, for
 1569 the subjective probabilities, mean $\Delta P = .51$ (SD = .46). 59 out of the 336 calculated ΔP s (17.6%)
 1570 were zero or negative and all came from the low correlation group. We also checked the best
 1571 fitting parameter values for the four rules in the implicit negations paradigm task and they also
 1572 all showed positive relevance (*if* p *then* q : $\Delta P = .43$; *if* p *then* $\neg q$: $\Delta P = .11$; *if* $\neg p$ *then* q : $\Delta P = .19$;

1573 *if* $\neg p$ *then* $\neg q$: $\Delta P = .09$). It would appear that for abstract conditionals (implicit negations
1574 paradigm) and those used in these experiments, people assume positive relevance between
1575 antecedent and consequent.

1576 Our results are relevant to an ongoing debate over the truth or acceptability conditions of
1577 conditionals. On the suppositional view of the conditional, judging whether a conditional is true
1578 or acceptable should depend on the conditional probability. According to semantic inferentialism
1579 (Douven, et al., 2018), in addition people must believe that there is an inferential link between
1580 antecedent and consequent. The existence of this inferential link explains why the antecedent
1581 *explains* the consequent for *if you turn the key the car starts*, but the antecedent of *if the moon is*
1582 *made of cheese, Corbyn will be elected Prime Minister* does not explain the consequent. Another
1583 example is the contrast between *if the sun rises, then the cock crows* and *if the cock crows then*
1584 *the sun rises*. Only in the former does the antecedent explain the consequent.¹⁷ This hypothesis
1585 has been tested by asking people how well the antecedent of an abductive or diagnostic
1586 conditional (e.g, *if the cock crows then the sun rises*) is explained by its consequent (Mirabile &
1587 Douven, in press: Experiment 3), thereby providing a measure of explanation quality.
1588 Participants also judged how strongly they believed the truth of the conclusion of an MP
1589 inference using the same abductive conditionals. Finally, they completed a probabilistic truth
1590 table task to obtain a measure of conditional probability. Explanation quality was a better
1591 predictor of how strongly someone believed that the conclusion of the MP inference was true
1592 than conditional probability. Explanation quality and conditional probability were also
1593 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).

1594 In looking at the relation between confidence and inference endorsement in the last
1595 section, we interpreted the fact that calculated conditional probability and confidence were
1596 highly correlated as indicating that confidence provided a more direct measure of conditional
1597 probability. That was why confidence was a better predictor of inference endorsement. The same
1598 argument applies to Mirabile and Douven's (in press; see also, Douven & Mirabile, 2018)
1599 measure of explanatory goodness, which they also assessed directly for each conditional.
1600 Consequently, explanatory goodness and confidence may just be better more direct measures of
1601 conditional probability than the probabilistic truth table task because they more closely follow
1602 the Ramsey test. So, contradicting Mirabile and Douven (in press), a construct of explanatory
1603 goodness distinct from conditional probability may not be required to explain the data.

1604 However, although this is a plausible line of argument, we would suggest that when you
1605 believe a conditional you believe it describes some underlying, usually causal, dependency in the
1606 world (Oaksford & Chater, 2010, 2017, 2020a, 2020b), which is why we suggested modelling
1607 these data using causal Bayes nets may be a fruitful line of research. That ΔP was positive for the
1608 main conditionals in our experiments showed that people believed the antecedent was positively
1609 causally relevant to the consequent because ΔP is the numerator of causal power (Cheng, 1997),
1610 which provides the weights on the links in a CBN (see *Supplementary Online Material*).
1611 Consequently, like semantic inferentialism, we would argue that the reason why confidence and
1612 explanation quality are better predictors of the odds of endorsing an inference is that people
1613 directly consider the causal or inferential link, which they do not need to do in the probabilistic
1614 truth table task. Indeed, if they learn a Bayes net during the learning phase, which requires them
1615 to consider the inferential link and its direction, then it would be difficult to reconstruct the
1616 individual cell values of the JPD in the probability verification task. It would require recording

1617 the prior over p , instantiating p to each of p_{1-3} and reading off the nine conditional probabilities
1618 $\Pr(q = q_{1-3} | p = p_{1-3})$ and multiplying them by the priors $\Pr(p = p_{1-3})$. That people seem capable
1619 of doing something like this with some degree of accuracy in the probability verification task is
1620 quite impressive. However, we learn about the world in order to predict and explain it and we
1621 argue that this requires setting up mental representations that facilitate inference, like the Bayes
1622 net in Figure 7.

1623

1624 **Learning**

1625 Our probability manipulations used brief experiential learning phases, shown in research in
1626 judgement and decision making to improve performance (Hogarth & Soyer, 2011; Wulf, et al.,
1627 2018). It is worth emphasizing that these learning experiences were short, only 30 trials in
1628 Experiments 2 and 3, and no attempt was made to get participants to learn the distributions to
1629 any criterion of accuracy. Nonetheless, these learning experiences profoundly influenced
1630 participants' behavior when presented with verbal conditional inference problems. All other
1631 theories attribute the implicit negations effect to errors in constructing a mental representation of
1632 the logical form of the premises. In contrast, we have argued that conditionals describe the
1633 dependencies in the world that allow us to predict and explain it (e.g., Oaksford & Chater, 2010,
1634 2020b). It should not be surprising that people are adept at rapidly acquiring the information they
1635 need from their immediate environment to build small scale models that allow them to do this
1636 and so to act in that environment.

1637 The importance of sampling from the environment is also emphasized in decision by
1638 sampling models (Sanborn & Chater, 2016; Stewart, et al. 2006). Samples may be derived from
1639 memory, but in novel contexts, where previous experience is little guide, people must sample

1640 from the environment. Moreover, the structure of samples or choice options can strongly
1641 influence decision making (Stewart, Chater, Stott, & Reimers, 2003). Models like Bayes nets,
1642 include information about structure (directed links and independence relations) and strength
1643 (causal strength or the relevant CPT). The probabilities that are used to compute strength can
1644 come from memory or, in novel contexts, must be sampled from the immediate environment. In
1645 Bayes nets there also are algorithms for learning not just the relevant probabilities but also the
1646 network structure of these models (Korb & Nicolson, 2010). That is, learning is integral to these
1647 models, in a way that it is not in other non-probabilistic theories of verbal reasoning. Moreover,
1648 as we have seen, how well participants learned the distribution strongly moderated the effect of
1649 calculated conditional probability and confidence on the odds of endorsing as inference.

1650 It could be argued that the reliance of our account, and its implementation in Bays nets,
1651 on learning is a limitation as it only applies when probabilities are learned. However, we have
1652 shown that the contrast set model also fits the base-line implicit negation effect (see, *Modelling*
1653 *the Default Prior Pr_0*). So the same model applies whether the probabilities are provided by
1654 memory or learned from the immediate environment. Although, of course, the default prior was
1655 also, presumably, learned, at least in part, from experience. Other probabilistic manipulations
1656 may be less effective in producing the discriminatory effects we observed in these experiments.
1657 So, Experiment 1 only showed minimal changes to the default prior when participants were
1658 given descriptions of the distribution in Table 2 as single event probabilities (e.g., 0.8 or 80%) in
1659 the pre-learning inference task. Single event probabilities, it would appear, do not update
1660 people's default-priors as effectively as experience, as many have argued (e.g., Gigerenzer &
1661 Hoffrage, 1995). However, it remains to be seen if frequency formats (80 out of a 100)
1662 (Gigerenzer & Hoffrage, 1995), lead to a more effective update as observed in some previous

1663 research (Oaksford, et al., 1997, 1999). Sample summaries (Hawkins et al., 2015) are closely
 1664 related to frequency formats. It would be interesting to see whether sample summaries of the
 1665 parameters of the CPT in Table 12 could produce similar effects. These distributions are the most
 1666 relevant to inference but they relate directly only to the forward inferences (MP and DA). An
 1667 interesting prediction of the Bayes net implementation is that when presented with only these
 1668 samples, the backwards inferences (AC and MT) should still track the inverse conditional
 1669 probabilities.

1670

1671 **Rationality**

1672 Is people's behavior on these tasks rational? Answering this question depends on what you think
 1673 people should do when confronted with these inference tasks. Clearly, people are not rational
 1674 with respect to standard conditional logic. Regardless of whether the negation in the
 1675 categorical premise is explicit or implicit, all that is logically relevant is whether it affirms or
 1676 denies the antecedent or consequent. If it affirms the antecedent (MP) or denies the consequent
 1677 (MT), the inference should be endorsed otherwise it should not be endorsed. Clearly, people are
 1678 not rational with respect to this standard as they happily reject inferences when a clause is denied
 1679 (affirmed) implicitly that they happily accept when it is denied (affirmed) explicitly.

1680 People can *deduce* probabilistic conclusions from uncertain premises (Cruz, Baratgin,
 1681 Oaksford, & Over, 2015; Evans, Thompson, & Over, 2015; Pfeifer & Kleiter, 2009; Politzer &
 1682 Baratgin, 2016; Singmann, Klauer, & Over, 2014). In coherence-based probability logics (Coletti
 1683 & Scozzafava, 2002), we can deduce a probability interval from the probabilities of the major
 1684 and minor premise. So, for example, suppose that in Experiments 1 and 2 $\Pr(\neg q|\neg p) = 0.8$ and
 1685 $\Pr(\neg p) = .8$, then the probability of the conclusion of MP must lie in the interval $.64 \geq \Pr(\neg q)$

1686 $\leq .84$. These intervals respect probabilistic coherence assuming only the information given in the
1687 premises. From this probabilistic logic point of view, again the only significance an implicit
1688 negation has is being an instance of the relevant negated category. In this paper, we have
1689 interpreted the evidence given by the categorical premise as either hard (affirmative) or virtual
1690 (negations) evidence concerning the states of the random variables in a Bayes net, which
1691 includes full knowledge of the JPD. Probability logic does not typically assume full knowledge
1692 of the JPD but allows for uncertainty in the categorical premise. Take for example AC, and
1693 assume that the probability of each categorical premise is the relevant marginal probability in
1694 Table 2. According to probabilistic coherence, for the explicit negation (AC-Not) the probability
1695 of the conclusion of this inference on (1) should be in the interval $[0, .278]$ and for implicit
1696 negation (AC-Con) it should be $[0, .937]$. However, the mean computed conditional probabilities
1697 and probabilities of endorsement (in brackets) of each inference was AC-Not: .79 (.97) and AC-
1698 Con: .77 (.94). For AC-Not both probabilities fell well outside of the coherence interval.
1699 Consequently, people's behavior in these experiments is not rational with respect to the standards
1700 of coherence-based probability logic.¹⁸

1701 From our perspective, reasoning is about rational change of belief (Eva & Hartmann,
1702 2018; Harman, 1986; Oaksford & Chater, 2007, 2020a). Here we have modelled inference as
1703 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).

1704 extent to which the relevant conditional probabilities predict inference endorsements show the
1705 extent to which we can view peoples' reasoning as rational. In our experimental tasks, the
1706 learning samples were taken from the same population of experiences as the informant (e.g., the
1707 vet) asserting the conditional, so the premises should not lead to any changes in the probabilities
1708 that define people's enduring beliefs in the CPT of their Bayes net representation.. However,
1709 there are situations where learning the premises suggests revisions to our degree of belief in a
1710 conditional premise (Oaksford & Chater, 2007, 2013). Such situations seem to require revising
1711 our beliefs not just updating them supposing the categorical premise is true. Although beyond the
1712 scope of our current discussion, guaranteeing the rationality of inference in these dynamic
1713 contexts remains a more challenging problem (Douven & Romeijn, 2011; Eva & Hartmann,
1714 2018, Hartmann & Rafiee Rad, 2012; Oaksford & Chater, 2013).

1715

1716 **Common Mechanisms**

1717 In explaining our results, we have not appealed to any mechanisms that are unique to deductive
1718 reasoning. Rather we have argued that mechanisms like Bayes nets may provide an account of
1719 the representations and processes underlying the implicit negation effect by providing an
1720 implementation of how people learn, represent and access contrast sets. We have previously
1721 argued that CBNs may provide an account of conditional inference, not just with causal
1722 conditionals (Ali et al., 2011), but with conditionals generally (Oaksford & Chater, 2010a,b). We
1723 have also argued that they may provide an implementation of inferentialism (Oaksford & Chater,
1724 2020b). More generally, we have argued that common mechanisms may underlie, inductive,
1725 deductive and causal reasoning and these are likely to be similar in kind to those that underlie
1726 judgement and decision-making (Oaksford & Chater, 2020a). Proposals for closer relations

1727 between deductive inference and other areas of higher cognition are not new: with judgement
1728 and decision-making (Manktelow & Over, 1991) and with causal reasoning (Oaksford & Chater,
1729 1994).

1730 However, there is a contrast with the mental models approach, which also provides
1731 explanations of inductive, deductive, and causal reasoning (Johnson-Laird, Goodwin, &
1732 Khemlani, 2018; Johnson-Laird, & Khemlani, 2017). Mental models treats discrete
1733 representations of possibilities as basic. These possibilities are closely related to the truth table
1734 cases allowed by the binary logical connectives, but they can be modulated by prior knowledge
1735 or labelled to capture other forms of inference. Following many other areas of perception and
1736 cognition, we regard the mind/brain's task to be the extraction of useful regularities from the flux
1737 of experience in order to predict and ultimately explain the world. The fundamental mode of
1738 representation is probabilistic and continuous, and it is only by sampling the brain's underlying
1739 stochastic models that we come to represent discrete possibilities. Usually these are just the
1740 deliverances to consciousness of the results of the processes that actually drive our behavior. If
1741 we do anything more with them it seems as likely to lead to error as to successful reasoning. So,
1742 while there is agreement on common mechanism, the new paradigm in reasoning generalizes in
1743 the opposite direction to mental models, from other areas of cognition to deduction and not from
1744 accounts of deductive reasoning elsewhere.

1745

1746

Conclusion

1747 Psychologists are beginning to uncover the rational basis for many of the biases discovered over
1748 the last 50 years in deductive and causal reasoning, judgement and decision-making. In this
1749 paper, we have argued that using a manipulation, experiential learning, shown to be effective in

1750 judgement and decision-making may elucidate the rational underpinning of the implicit negation
1751 effect in conditional inference. In three experiments, we created and removed the effect by using
1752 probabilistically structured contrast sets acquired during a brief learning phase. No other theory
1753 of the implicit negations effect makes these predictions. We could model our findings well using
1754 Bayes nets similar to causal approaches to category structure, which also captured further
1755 intuitions about how contrast sets can identify the most likely opposites. We also showed that our
1756 results and our Bayes net approach aligns closely to a recent development in the psychology of
1757 reasoning called inferentialism. A key feature is that we have not appealed to any cognitive
1758 mechanism or module whose specific task is logical reasoning. This approach is consistent with
1759 the conclusion of our recent review of new paradigm probabilistic theories, which treats
1760 argumentation, deduction and induction alike within a probabilistic framework similar in kind to
1761 processes involved in other areas of cognition (Oaksford & Chater, 2020a).

1762

1763

Context

1764 We have been explaining biases in human deductive reasoning using Bayesian rational analysis
1765 for 25 years (Oaksford & Chater, 1994, 2020a). This pattern of explanation had seemed
1766 paradoxical because Bayesian reasoning in judgement and decision-making had always seemed
1767 similarly biased. Recently, however, it has been shown that people's judgement and decision-
1768 making can be surprisingly rational when probabilities and utilities are learned by experience.
1769 We used experiential learning phases to allow participants to acquire information about
1770 probability distributions that should create and remove the implicit negation effect in conditional
1771 reasoning. This is the first time that discrete experiential learning has been used to manipulate
1772 probabilities in deductive reasoning tasks. We had already shown that our Bayesian approach

1773 could rationally explain polarity biases in conditional inference using the concept of a contrast
1774 set. Our current experiments show that this account generalises to the implicit negations effect.
1775 We could also model the effects well using Bayes nets. We show how these data also apply
1776 directly to recent inferentialist accounts of conditional inference. Our results suggest that similar
1777 cognitive mechanisms may underlie causal, inductive and deductive reasoning as proposed in our
1778 recent review of the new paradigm in the psychology of reasoning (Oaksford & Chater, 2020a).

1779 References

- 1780 Ali, N., Chater, N., & Oaksford, M. (2011). The mental representation of causal conditional
1781 inference: Causal models or mental models. *Cognition*, *119*, 403-418.
- 1782 Ali, N, Schlottmann, A., Shaw, C., Chater, N., & Oaksford, M., (2010). Conditionals and causal
1783 discounting in children. In M. Oaksford & N. Chater (Eds.), *Cognition and conditionals:
1784 Probability and logic in human thinking* (pp. 117-134). Oxford: Oxford University Press.
- 1785 Anderson, J. R. (1995). *Cognitive psychology and its implications*. New York: W. H. Freeman.
- 1786 Anderson, J. R., & Sheu, C.-F. (1995). Causal inferences as perceptual judgements. *Memory &
1787 Cognition*, *23*, 510–524. <https://doi-org.ezproxy.lib.bbk.ac.uk/10.3758/BF03197251>
- 1788 Anderson, S. K., Olesen, K. G., & Jensen, F. V. (1990). HUGIN—a shell for building Bayesian
1789 belief universes for expert systems. In G. Shafer & J. Pearl (Eds.), *Readings in uncertain
1790 reasoning* (pp. 332-337). San Francisco, CA: Morgan Kaufmann Publishers Inc.
- 1791 Ardia, D., Mullen, K. M., Peterson, B. G., & Ulrich, J. (2016). DEoptim: Differential Evolution
1792 in 'R'. version 2.2-4.
- 1793 Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed
1794 random effects for subjects and items. *Journal of Memory and Language*, *59*, 340–412.
1795 <http://dx.doi.org/10.1016/j.jml.2007.12.005>
- 1796 Bakdash, J. Z., & Marusich, L. R. (2017). Repeated measures correlation. *Frontiers in
1797 Psychology*, *8*, 456. doi: 10.3389/fpsyg.2017.00456
- 1798 Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition*, *11*, 211–227. [https://doi-
1799 org.ezproxy.lib.bbk.ac.uk/10.3758/BF03196968](https://doi-org.ezproxy.lib.bbk.ac.uk/10.3758/BF03196968)

- 1800 Barsalou, L. W., Huttenlocher, J., & Lamberts, K. (1998). Basing categorization on individuals
1801 and events. *Cognitive Psychology*, *36*, 203–272. [https://doi-](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1006/cogp.1998.0687)
1802 [org.ezproxy.lib.bbk.ac.uk/10.1006/cogp.1998.0687](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1006/cogp.1998.0687)
- 1803 Bilmes, J. (2004). *On virtual evidence and soft evidence in Bayesian networks* (UWEE Technical
1804 Report No. UWEETR-2004-0016). Seattle, Washington: Department of Electrical
1805 Engineering.
- 1806 Byrne, R. M. J., & Johnson-Laird, P. N. (2019). If and or: Real and counterfactual possibilities in
1807 their truth and probability. *Journal of Experimental Psychology: Learning, Memory, and*
1808 *Cognition*. <https://doi-org.ezproxy.lib.bbk.ac.uk/10.1037/xlm0000756.supp>
1809 (Supplemental)
- 1810 Chater, N. (2018). *The mind is flat*. London: Allen Lane.
- 1811 Chater, N., & Oaksford, M. (2006). Mental mechanisms. In K. Fiedler, & P. Juslin (Eds.),
1812 *Information sampling and adaptive cognition* (pp. 210-236). Cambridge: Cambridge
1813 University Press.
- 1814 Cheng, P. (1997). From covariation to causation: A causal power theory. *Psychological Review*,
1815 *104*, 367–405.
- 1816 Chow S, Shao J, Wang H. (2008). *Sample Size Calculations in Clinical Research*. 2nd Ed.
1817 Chapman & Hall/CRC Biostatistics Series
- 1818 Coletti, G., & Scozzafava, R. (2002). *Probabilistic logic in a coherent setting*. Dordrecht, NL:
1819 Kluwer.
- 1820 Costello, F., & Watts, P. (2014). Surprisingly rational: Probability theory plus noise explains
1821 biases in judgement. *Psychological Review*, *121*, 463–480. [https://doi-](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1037/a0037010)
1822 [org.ezproxy.lib.bbk.ac.uk/10.1037/a0037010](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1037/a0037010)

- 1823 Costello, F., Watts, P., & Fisher, C. (2018). Surprising rationality in probability judgement:
1824 Assessing two competing models. *Cognition*, *170*, 280–297. [https://doi-](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1016/j.cognition.2017.08.012)
1825 [org.ezproxy.lib.bbk.ac.uk/10.1016/j.cognition.2017.08.012](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1016/j.cognition.2017.08.012)
- 1826 Cruz, N., Baratgin, J., Oaksford, M., & Over, D. E. (2015). Reasoning with if and ands and ors.
1827 *Frontiers in Psychology*, *6*, 192. doi: 10.3389/fpsyg.2015.00192
- 1828 Dasgupta, I., Schulz, E., & Gershman, S. J. (2017). Where do hypotheses come from? *Cognitive*
1829 *Psychology*, *96*, 1–25. [https://doi-](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1016/j.cogpsych.2017.05.001)
1830 [org.ezproxy.lib.bbk.ac.uk/10.1016/j.cogpsych.2017.05.001](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1016/j.cogpsych.2017.05.001)
- 1831 Davis, Z., & Rehder, B. (2017). The causal sampler: A sampling approach to causal
1832 representation, reasoning and learning. *Poster presented at the 39th annual conference of*
1833 *the Cognitive Science Society*. London, UK.
- 1834 De Martino, B., Fleming, S. M., Garrett, N., & Dolan, R. J. (2013). Confidence in value-based
1835 choice. *Nature Neuroscience*, *16*, 105–110. [https://doi-](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1038/nn.3279)
1836 [org.ezproxy.lib.bbk.ac.uk/10.1038/nn.3279](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1038/nn.3279)
- 1837 Douven, I., Elqayam, S., Singmann, H., & van Wijnbergen-Huitink, J. (2018). Conditionals and
1838 inferential connections: A hypothetical inferential theory. *Cognitive Psychology*, *101*, 50–
1839 81. <https://doi-org.ezproxy.lib.bbk.ac.uk/10.1016/j.cogpsych.2017.09.002>
- 1840 Douven, I., & Mirabile, P. (2018). Best, second-best, and good-enough explanations: How they
1841 matter to reasoning. *Journal of Experimental Psychology: Learning, Memory, and*
1842 *Cognition*, *44*, 1792–1813. <https://doi-org.ezproxy.lib.bbk.ac.uk/10.1037/xlm0000545>
- 1843 Douven, I., & Romeijn, J.-W. (2011). A new resolution of the Judy Benjamin problem. *Mind*,
1844 *120*, 637–670.

- 1845 Edgington, D. (1995). On conditionals. *Mind*, *104*, 235–329.
- 1846 <http://dx.doi.org/10.1093/mind/104.414.235>
- 1847 Eva, B., & Hartmann, S. (2018). Bayesian argumentation and the value of logical
1848 validity. *Psychological Review*, *125*, 806–821. [https://doi-](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1037/rev0000114.supp)
1849 [org.ezproxy.lib.bbk.ac.uk/10.1037/rev0000114.supp](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1037/rev0000114.supp) (Supplemental)
- 1850 Evans, J. St.B. T. (1998). Matching bias in conditional reasoning: Do we understand it after 25
1851 years? *Thinking and Reasoning*, *4*, 45–82.
- 1852 Evans, J.St.B.T. (2002). Matching bias and set sizes: A discussion of Yama (2001). *Thinking &*
1853 *Reasoning*, *8*, 153-163.
- 1854 Evans, J.St.B.T. (1972). Interpretation and “matching bias” in a reasoning task. *Quarterly*
1855 *Journal of Experimental Psychology*, *24*, 193–199.
- 1856 Evans, J. St.B. T., Clibbens, J., & Rood, B. (1995). Bias in conditional inference: Implications
1857 for mental models and mental logic. *Quarterly Journal of Experimental Psychology*, *48A*,
1858 644–670. <https://doi-org.ezproxy.lib.bbk.ac.uk/10.1080/14640749508401409>
- 1859 Evans, J.St.B.T., Clibbens, J., & Rood, B. (1996). The role of implicit and explicit negations in
1860 conditional reasoning bias. *Journal of Memory & Language*, *35*, 392–409.
- 1861 Evans, J.St.B.T., & Handley, S.J. (1999). The role of negation in conditional inference.
1862 *Quarterly Journal of Experimental Psychology*, *52*, 739–770.
- 1863 Evans, J. St. B. T., Handley, S. J., & Over, D. E. (2003). Conditionals and conditional
1864 probability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*,
1865 321–335. <http://dx.doi.org/10.1037/0278-7393.29.2.321>
- 1866 Evans, J.St.B.T., & Lynch, J.S. (1973). Matching bias in the selection task. *British Journal of*
1867 *Psychology*, *64*, 391–397.

- 1868 Evans, J. St.B. T., Thompson, V., & Over, D. E. (2015). Uncertain deduction and conditional
1869 reasoning. *Frontiers in Psychology*, 6, 398.
- 1870 Fernbach, P. M., & Erb, C. D. (2013). A quantitative causal model theory of conditional
1871 reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39,
1872 1327-1343. doi:10.1037/a0031851
- 1873 Fugard, A. J. B., Pfeifer, N., Mayerhofer, B., & Kleiter, G. D. (2011). How people interpret
1874 conditionals: Shifts toward the conditional event. *Journal of Experimental Psychology:
1875 Learning, Memory, and Cognition*, 37, 635–648. [https://doi-](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1037/a0022329)
1876 [org.ezproxy.lib.bbk.ac.uk/10.1037/a0022329](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1037/a0022329)
- 1877 Gelman, A. (2018). You need 16 times the sample size to estimate an interaction than to estimate
1878 a main effect. Retrieved from: [https://statmodeling.stat.columbia.edu/2018/03/15/need-](https://statmodeling.stat.columbia.edu/2018/03/15/need-16-times-sample-size-estimate-interaction-estimate-main-effect/)
1879 [16-times-sample-size-estimate-interaction-estimate-main-effect/](https://statmodeling.stat.columbia.edu/2018/03/15/need-16-times-sample-size-estimate-interaction-estimate-main-effect/)
- 1880 Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013).
1881 *Bayesian data analysis*. New York: CRC Press.
- 1882 Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction:
1883 Frequency formats. *Psychological Review*, 102, 684-704.
- 1884 Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive
1885 Psychology*, 51, 334–384.
- 1886 Goodrich, B., Gabry, J., Ali, I., Brilleman, S. (2018). rstanarm: Bayesian applied regression
1887 modeling via Stan. R package version 2.17.4, <http://mc-stan.org/>.
- 1888 Hahn, U., & Oaksford, M. (2007). The rationality of informal argumentation: A Bayesian
1889 approach to reasoning fallacies. *Psychological Review*, 114, 704-732.
- 1890 Harman, G. (1986). *Change in view: Principles of reasoning*. Cambridge, MA: MIT

- 1891 Press/Bradford Books.
- 1892 Hartmann, S., & Rafiee Rad, S. (2012). Updating on conditionals = Kullback-Leibler distance +
1893 causal structure. Presented at the *Biennial Meeting of the Philosophy of Science*
1894 *Association*, San Diego, CA, November 15–17.
- 1895 Hattori, M. (2016). Probabilistic representation in syllogistic reasoning: A theory to integrate
1896 mental models and heuristics. *Cognition*, *157*, 296–320. [https://doi-](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1016/j.cognition.2016.09.009)
1897 [org.ezproxy.lib.bbk.ac.uk/10.1016/j.cognition.2016.09.009](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1016/j.cognition.2016.09.009)
- 1898 Hattori, M., & Oaksford, M. (2007). Adaptive non-interventional heuristics for co-variation
1899 detection in causal induction: Model comparison and rational analysis. *Cognitive Science*,
1900 *31*, 765-814.
- 1901 Hawkins, G. E., Hayes, B. K., Donkin, C., Pasqualino, M., & Newell, B. R. (2015). A Bayesian
1902 latent-mixture model analysis shows that informative samples reduce base-rate
1903 neglect. *Decision*, *2*, 306–318. [https://doi-](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1037/dec0000024.supp)
1904 [org.ezproxy.lib.bbk.ac.uk/10.1037/dec0000024.supp](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1037/dec0000024.supp) (Supplemental)
- 1905 Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the
1906 effect of rare events in risky choice. *Psychological Science*, *15*, 534 –539.
1907 <http://dx.doi.org/10.1111/j.0956-7976.2004.00715.x>
- 1908 Hinterecker, T., Knauff, M., & Johnson-Laird, P. N. (2016). Modality, probability, and mental
1909 models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*,
1910 1606–1620. <http://dx.doi.org/10.1037/xlm0000255>
- 1911 Hogarth, R. M., & Soyer, E. (2011). Sequentially simulated outcomes: Kind experience versus
1912 non-transparent description. *Journal of Experimental Psychology: General*, *140*, 434.
1913 <http://dx.doi.org/10.1037/a0023265>

- 1914 Horn, L. R. (1989). *A natural history of negation*. Chicago: University of Chicago Press.
- 1915 Jarvstad, A., Hahn, U., Rushton, S. K., & Warren, P. A. (2013). Perceptuo-motor, cognitive, and
1916 description-based decision-making seem equally good. *PNAS Proceedings of the*
1917 *National Academy of Sciences of the United States of America*, *110*, 16271–16276.
1918 <https://doi-org.ezproxy.lib.bbk.ac.uk/10.1073/pnas.1300239110>
- 1919 Johnson-Laird, P. N. (1983). *Mental models*. Cambridge, UK; Cambridge University Press
- 1920 Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics,
1921 and inference. *Psychological Review*, *109*, 646–678. [https://doi-](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1037/0033-295X.109.4.646)
1922 [org.ezproxy.lib.bbk.ac.uk/10.1037/0033-295X.109.4.646](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1037/0033-295X.109.4.646)
- 1923 Johnson-Laird, P. N., Goodwin, G. P., & Khemlani, S. S. (2018). Mental models and reasoning.
1924 In L. J. Ball & V. A. Thompson (Eds.), *The Routledge international handbook of thinking*
1925 *and reasoning*. (pp. 346–365). New York, NY: Routledge/Taylor & Francis Group.
- 1926 Johnson-Laird, P. N., & Khemlani, S. S. (2017). Mental models and causation. In M. R.
1927 Waldmann (Ed.), *The Oxford handbook of causal reasoning*. (pp. 169–187). New York,
1928 NY: Oxford University Press.
- 1929 Johnson-Laird, P. N., Legrenzi, P., Girotto, V., Legrenzi, M. S., & Caverni, J. P. (1999). Naïve
1930 probability: A mental model theory of extensional reasoning. *Psychological Review*, *106*,
1931 62–88.
- 1932 Joseph, L., du Berger, R., & Belisle P. (1997). Bayesian and mixed Bayesian/likelihood criteria
1933 for sample size determination propdiff.acc 3. *Statistics in Medicine*, *16*, 769-781.
- 1934 Jubin, J., & Barrouillet, P. (2019). Effects of context on the rate of conjunctive responses in the
1935 probabilistic truth table task. *Thinking & Reasoning*, *25*, 133–150.
1936 <https://doi.org/10.1080/13546783.2018.1477689>

- 1937 Kay, M. (2019). tidybayes: Tidy data and geoms for Bayesian models. doi:
1938 10.5281/zenodo.1308151 (URL: <http://doi.org/10.5281/zenodo.1308151>), R package
1939 version 1.0.4, <URL: <http://mjskay.github.io/tidybayes/>>.
- 1940 Khemlani, S., Hinterecker, T., & Johnson-Laird, P. N. (2017). The provenance of modal
1941 inference. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. J. Davelaar (Eds.),
1942 *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (pp. 663–
1943 668). Austin, TX: Cognitive Science Society.
- 1944 Khemlani, S., Orenes, I., & Johnson-Laird, P. N. (2012). Negation: A theory of its meaning,
1945 representation, and use. *Journal of Cognitive Psychology*, 24, 541–559. [https://doi-](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1080/20445911.2012.660913)
1946 [org.ezproxy.lib.bbk.ac.uk/10.1080/20445911.2012.660913](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1080/20445911.2012.660913)
- 1947 Kleiter, G. D., (1994). Natural sampling: Rationality without base rates. In G. H. Fischer & D.
1948 Laming (Eds.), *Contributions to mathematical psychology, psychometrics, and*
1949 *methodology* (pp. 375-388). New York: Springer.
- 1950 Korb, K. B., & Nicholson, A. E. (2010). *Bayesian artificial intelligence*. New York: CRC Press.
- 1951 Kruschke, J. (2011). *Doing Bayesian data analysis*. New York: Academic Press.
- 1952 Krzyżanowska, K., Wenmackers, S., & Douven, I. (2013). Inferential conditionals and
1953 evidentiality. *Journal of Logic, Language and Information*, 22, 315–334. [https://doi-](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1007/s10849-013-9178-4)
1954 [org.ezproxy.lib.bbk.ac.uk/10.1007/s10849-013-9178-4](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1007/s10849-013-9178-4)
- 1955 Lenth, R. (2019). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package
1956 version 1.3.3. <https://CRAN.R-project.org/package=emmeans>
- 1957 Lüdtke, D. (2018). sjPlot: Data visualization for statistics in social science. doi:
1958 10.5281/zenodo.1308157 (URL: <http://doi.org/10.5281/zenodo.1308157>), R package
1959 version 2.6.2, <URL: <https://CRAN.R-project.org/package=sjPlot>>.

- 1960 Marcus, S. L., & Rips, L. J. (1979). Conditional reasoning. *Journal of Verbal Learning & Verbal*
1961 *Behavior, 18*, 199–223. <https://doi-org.ezproxy.lib.bbk.ac.uk/10.1016/S0022->
1962 [5371\(79\)90127-0](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1016/S0022-5371(79)90127-0)
- 1963 McElreath, R. (2016). *Statistical rethinking*. New York: CRC Press.
- 1964 McKenzie, C. R. M. (2004). Framing effects in inference tasks and why they are normatively
1965 defensible. *Memory and Cognition, 32*, 874-885.
- 1966 McKenzie, C. R. M., Ferreira, V. S., Mikkelsen, L. A., McDermott, K. J., & Skrable, R. P.
1967 (2001). Do conditional statements target rare events? *Organizational Behavior & Human*
1968 *Decision Processes, 85*, 291-309.
- 1969 McKenzie, C. R. M., & Mikkelsen, L. A. (2000). The psychological side of Hempel’s paradox of
1970 confirmation. *Psychonomic Bulletin & Review, 7*, 360-366.
- 1971 McKenzie, C. R. M., & Mikkelsen, L. A. (2007). A Bayesian view of covariation assessment.
1972 *Cognitive Psychology, 54*, 33–61.
- 1973 Mirabile, P., & Douven, I. (in press). Abductive conditionals as a test case for inferentialism.
1974 *Cognition*.
- 1975 Mrad, A. B., Delcroix, V., Piechowiak, S., Leicester, P., Mohamed, A. (2015). An explication of
1976 uncertain evidence in Bayesian networks: likelihood evidence and probabilistic evidence.
1977 *Applied Intelligence, 43*, 802–824. <https://doi.org/10.1007/s10489-015-0678-6>
- 1978 Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship.
1979 *Journal of Experimental Psychology: General, 115*, 39-57.
- 1980 Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization
1981 of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and*
1982 *Cognition, 13*, 87-109.

- 1983 Oaksford, M. (2002). Contrast classes and matching bias as explanations of the effects of
1984 negation on conditional reasoning. *Thinking & Reasoning*, 8, 135-151.
- 1985 Oaksford, M. & Chater, N. (1994). A rational analysis of the selection task as optimal data
1986 selection. *Psychological Review*, 101, 608-631.
- 1987 Oaksford, M. & Chater, N. (1996). Rational explanation of the selection task. *Psychological*
1988 *Review*, 103, 381-391.
- 1989 Oaksford, M., & Chater, N. (2003a). Computational levels and conditional reasoning. *Journal of*
1990 *Experimental Psychology: Learning, Memory & Cognition*, 29, 150-156.
- 1991 Oaksford, M., & Chater, N. (2003b). Optimal data selection: Revision, review and re-evaluation.
1992 *Psychonomic Bulletin & Review*, 10, 289-318.
- 1993 Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human*
1994 *reasoning*. Oxford: Oxford University Press.
- 1995 Oaksford, M., & Chater, N. (2010a). Conditionals and constraint satisfaction: Reconciling
1996 mental models and the probabilistic approach? In M. Oaksford & N. Chater (Eds.),
1997 *Cognition and conditionals: Probability and logic in human thinking* (pp. 309-334).
1998 Oxford: Oxford University Press.
- 1999 Oaksford, M., & Chater, N. (2010b). Causation and conditionals in the cognitive science of
2000 human reasoning. *Open Psychology Journal*, 3, 105-118. (Special issue on *Causal learning*
2001 *beyond causal judgement*, Edited by J. C. Perales, & D. R. Shanks).
- 2002 Oaksford, M., & Chater, N. (2013). Dynamic inference and everyday conditional reasoning in
2003 the new paradigm. *Thinking and Reasoning*, 19, 346-379. doi:
2004 10.1080/13546783.2013.808163

- 2005 Oaksford, M., & Chater, N. (2016). Probabilities, causation, and logic programming in
2006 conditional reasoning: A reply to Stenning and van Lambalgen. *Thinking and Reasoning*,
2007 22, 336-354. doi: 10.1080/13546783.2016.1139505
- 2008 Oaksford, M., & Chater, N. (2017). Causal models and conditional reasoning. In M. Waldmann
2009 (Ed.), *Oxford handbook of causal reasoning* (pp. 327-346). Oxford: Oxford University
2010 Press.
- 2011 Oaksford, M., & Chater, N. (2020a). New paradigms in the psychology of reasoning. *Annual*
2012 *Review of Psychology*, 71, 305-330.
- 2013 Oaksford, M., & Chater, N. (2020b). Integrating Causal Bayes Nets and inferentialism in
2014 conditional inference. In S. Elqayam, I. Douven, J. St.B. T. Evans, & N. Cruz (Eds.). *Logic*
2015 *and uncertainty in the human mind: A tribute to David E. Over*. London: Routledge.
- 2016 Oaksford, M., Chater, N., & Grainger, B. (1999). Probabilistic effects in data selection. *Thinking*
2017 *and Reasoning*, 5, 193–243.
- 2018 Oaksford, M., Chater, N., Grainger, B., & Larkin, J. (1997). Optimal data selection in the
2019 reduced array selection task (RAST). *Journal of Experimental Psychology: Learning,*
2020 *Memory & Cognition*, 23, 441-458.
- 2021 Oaksford, M., Chater, N., & Larkin, J. (2000). Probabilities and polarity biases in conditional
2022 inference. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 26, 883-
2023 899.
- 2024 Oaksford, M., & Moussakowski, M. (2004). Negations and natural sampling in data selection:
2025 Ecological vs. heuristic explanations of matching bias. *Memory & Cognition*, 32, 570-581.

- 2026 Oaksford, M., Over, D. E., & Cruz, N. (2019). Paradigms, possibilities and probabilities:
2027 Comment on Hinterecker, Knauff, and Johnson-Laird (2016). *Journal of Experimental*
2028 *Psychology: Learning, Memory, & Cognition*, *45*, 288-297.
- 2029 Oaksford, M., & Stenning, K. (1992). Reasoning with conditionals containing negated
2030 constituents. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *18*,
2031 835–854.
- 2032 Oaksford, M., & Wakefield, M. (2003). Data selection and natural sampling: Probabilities do
2033 matter. *Memory & Cognition*, *31*, 143-154.
- 2034 Oberauer, K. (2006). Reasoning with conditionals: A test of formal models of four
2035 theories. *Cognitive Psychology*, *53*, 238–283. [https://doi-](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1016/j.cogpsych.2006.04.001)
2036 [org.ezproxy.lib.bbk.ac.uk/10.1016/j.cogpsych.2006.04.001](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1016/j.cogpsych.2006.04.001)
- 2037 Oberauer, K., & Wilhelm, O. (2003). The meaning(s) of conditionals: Conditional probabilities,
2038 mental models, and personal utilities. *Journal of Experimental Psychology: Learning,*
2039 *Memory, and Cognition*, *29*, 680–693. <http://dx.doi.org/10.1037/0278-7393.29.4.680>
- 2040 Oberauer, K., Weidenfeld, A., & Hörnig, R. (2004). Logical reasoning and probabilities: A
2041 comprehensive test of Oaksford and Chater (2001). *Psychonomic Bulletin & Review*, *11*,
2042 521–527. <https://doi-org.ezproxy.lib.bbk.ac.uk/10.3758/BF03196605>
- 2043 Orenes, I., Beltrán, D., & Santamaría, C. (2014). How negation is understood: Evidence from the
2044 visual world paradigm. *Journal of Memory and Language*, *74*, 36–45. [https://doi-](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1016/j.jml.2014.04.001)
2045 [org.ezproxy.lib.bbk.ac.uk/10.1016/j.jml.2014.04.001](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1016/j.jml.2014.04.001)
- 2046 Over, D. E., Hadjichristidis, C., Evans, J. S. B. T., Handley, S. J., & Sloman, S. A. (2007). The
2047 probability of causal conditionals. *Cognitive Psychology*, *54*, 62–97. [https://doi-](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1016/j.cogpsych.2006.05.002)
2048 [org.ezproxy.lib.bbk.ac.uk/10.1016/j.cogpsych.2006.05.002](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1016/j.cogpsych.2006.05.002)

- 2049 Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. San Mateo: Morgan Kaufmann.
- 2050 Pfeifer, N., & Kleiter, G. D. (2009). Framing human inference by coherence based probability
2051 logic. *Journal of Applied Logic*, 7, 206-217.
- 2052 Politzer, G., & Baratgin, J. (2016). Deductive schemas with uncertain premises using qualitative
2053 probability expressions. *Thinking & Reasoning*, 22, 78-98.
- 2054 Pollard, P., & Evans, J. St.B. T. (1983). The effect of experimentally contrived experience on
2055 reasoning performance. *Psychological Research*, 45, 287–301. [https://doi-](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1007/BF00308708)
2056 [org.ezproxy.lib.bbk.ac.uk/10.1007/BF00308708](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1007/BF00308708)
- 2057 Pothos, E. M., & Busemeyer, J. R. (2013). Can quantum probability provide a new direction for
2058 cognitive modeling? *Behavioral and Brain Sciences*, 36, 255–274. [https://doi-](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1017/S0140525X12001525)
2059 [org.ezproxy.lib.bbk.ac.uk/10.1017/S0140525X12001525](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1017/S0140525X12001525)
- 2060 Pothos, E. M., Busemeyer, J. R., Shiffrin, R. M., & Yearsley, J. M. (2017). The rational status of
2061 quantum cognition. *Journal of Experimental Psychology: General*, 146, 968–987.
2062 <https://doi-org.ezproxy.lib.bbk.ac.uk/10.1037/xge0000312>
- 2063 R Core Team (2018). R: A language and environment for statistical computing. R Foundation for
2064 Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- 2065 Rehder, B. (2003a). A causal-model theory of conceptual representation and categorization.
2066 *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 1141–1159.
- 2067 Rehder, B. (2003b). Categorization as causal reasoning. *Cognitive Science*, 27, 709–748.
- 2068 Rehder, B. (2014). Independence and dependence in human causal reasoning. *Cognitive*
2069 *Psychology*, 72, 54–107.

- 2070 Rehder, B. (2017). Concepts as causal models: Classification. In M. R. Waldmann (Ed.). *The*
2071 *Oxford handbook of causal reasoning* (pp. 346–375). New York, NY: Oxford University
2072 Press.
- 2073 Rehder, B., (2018). Beyond Markov: Accounting for independence violations in causal
2074 reasoning, *Cognitive Psychology*, *103*, 42–84
- 2075 Rips, L. J. (1994). *The psychology of proof: Deductive reasoning in human thinking*. Cambridge,
2076 MA: The MIT Press.
- 2077 Rottman, B., & Hastie, R. (2016). Do people reason rationally about causally related events?
2078 Markov violations, weak inferences, and failures of explaining away. *Cognitive*
2079 *Psychology*, *87*, 88–134.
- 2080 Rozenblit, L. & Keil, F. (2002). The misunderstood limits of folk science: An illusion of
2081 explanatory depth. *Cognitive Science*, *26*, 521-562.
- 2082 Sanborn, A. N., & Chater, N. (2016). Bayesian brains without probabilities. *Trends in Cognitive*
2083 *Sciences*, *20*, 883–893.
- 2084 Schroyens, W., & Schaeken, W. (2003). A critique of Oaksford, Chater, and Larkin’s (2000)
2085 conditional probability model of conditional reasoning. *Journal of Experimental*
2086 *Psychology: Learning, Memory, and Cognition*, *29*, 140–149. [https://doi-](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1037/0278-7393.29.1.140)
2087 [org.ezproxy.lib.bbk.ac.uk/10.1037/0278-7393.29.1.140](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1037/0278-7393.29.1.140)
- 2088 Schroyens, W., Schaeken, W., & d’Ydewalle (2001). The processing of negations in conditional
2089 reasoning: A meta-analytic case study in mental logic and/or mental models. *Thinking and*
2090 *Reasoning*, *7*, 121–172.

- 2091 Schroyens, W., Schaeken, W., Fias, W., & d'Ydewalle (2000). Heuristic and analytic processes
2092 in propositional reasoning with negatives. *Journal of Experimental Psychology: Learning,*
2093 *Memory & Cognition, 26*, 1713–1734.
- 2094 Schroyens, W., Schaeken, W., Verschueren, N., & d'Ydewalle (2000). Conditional reasoning
2095 with negations: Implicit and explicit affirmation or denial and the role of contrast classes.
2096 *Thinking and Reasoning, 6*, 221–252.
- 2097 Schwan, B., & Stern, R. (2017). A causal understanding of when and when not to Jeffrey
2098 conditionalize. *Philosophers' Imprint, 17*, 8.
- 2099 Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of*
2100 *Mathematical Psychology, 1*, 54–87. [https://doi-org.ezproxy.lib.bbk.ac.uk/10.1016/0022-](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1016/0022-2496(64)90017-3)
2101 [2496\(64\)90017-3](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1016/0022-2496(64)90017-3)
- 2102 Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2019). afex: Analysis of Factorial
2103 Experiments. R package version 0.23-0. <https://CRAN.R-project.org/package=afex>
- 2104 Singmann, H., Klauer, K. C., & Over, D. E. (2014). New normative standards of conditional
2105 reasoning and the dual-source model. *Frontiers in Psychology, 5*, 316.
- 2106 Skovgaard-Olsen, N., Collins, P., Krzyżanowska, K., Hahn, U., & Klauer, K. C. (2019).
2107 Cancellation, negation, and rejection. *Cognitive Psychology, 108*, 42–71. [https://doi-](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1016/j.cogpsych.2018.11.002)
2108 [org.ezproxy.lib.bbk.ac.uk/10.1016/j.cogpsych.2018.11.002](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1016/j.cogpsych.2018.11.002)
- 2109 Skovgaard-Olsen, N., Kellen, D., Hahn, U., & Klauer, K. C. (2019). Norm conflicts and
2110 conditionals. *Psychological Review, 126*, 611–633. [https://doi-](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1037/rev0000150.supp)
2111 [org.ezproxy.lib.bbk.ac.uk/10.1037/rev0000150.supp](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1037/rev0000150.supp) (Supplemental)
- 2112 Skovgaard-Olsen, N., Kellen, D., Krahl, H., & Klauer, K. C. (2017). Relevance differently
2113 affects the truth, acceptability, and probability evaluations of 'and', 'but', 'therefore', and

- 2114 “if-then.” *Thinking & Reasoning*, 23, 449–482. <https://doi->
- 2115 [org.ezproxy.lib.bbk.ac.uk/10.1080/13546783.2017.1374306](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1080/13546783.2017.1374306)
- 2116 Skovgaard-Olsen, N., Singmann, H., & Klauer, K. C. (2016). The relevance effect and
- 2117 conditionals. *Cognition*, 150, 26–36. <https://doi->
- 2118 [org.ezproxy.lib.bbk.ac.uk/10.1016/j.cognition.2015.12.017](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1016/j.cognition.2015.12.017)
- 2119 Skovgaard-Olsen, N., Singmann, H., & Klauer, K. C. (2017). Relevance and reason
- 2120 relations. *Cognitive Science*, 41, 1202–1215. <https://doi->
- 2121 [org.ezproxy.lib.bbk.ac.uk/10.1111/cogs.12462](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1111/cogs.12462)
- 2122 Sloman, S. A., & Fernbach, P. (2017). *The knowledge illusion: Why we never think alone*. New
- 2123 York, NY: Riverhead books.
- 2124 Stahl, C., Klauer, K. C., & Erdfelder, E. (2008). Matching bias in the selection task is not
- 2125 eliminated by explicit negations. *Thinking & Reasoning*, 14, 281–303. <https://doi->
- 2126 [org.ezproxy.lib.bbk.ac.uk/10.1080/13546780802116807](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1080/13546780802116807)
- 2127 Stewart, N., Chater, N., & Brown, G. D. A. (2006). Decision by sampling. *Cognitive*
- 2128 *Psychology*, 53, 1–26. <https://doi->
- 2129 [org.ezproxy.lib.bbk.ac.uk/10.1016/j.cogpsych.2005.10.003](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1016/j.cogpsych.2005.10.003)
- 2130 Stewart, N., Chater, N., Stott, H. P., & Reimers, S. (2003). Prospect relativity: How choice
- 2131 options influence decision under risk. *Journal of Experimental Psychology: General*,
- 2132 132, 23–46.
- 2133 Tentori, K., Crupi, V., & Russo, S. (2013). On the determinants of the conjunction fallacy
- 2134 Probability versus inductive confirmation. *Journal of Experimental Psychology:*
- 2135 *General*, 142, 235–255. <https://doi-org.ezproxy.lib.bbk.ac.uk/10.1037/a0028770>

- 2136 Thompson, V. A., Evans, J. St.B. T., & Campbell, J. I. D. (2013). Matching bias on the selection
2137 task: It's fast and feels good. *Thinking & Reasoning*, 19(, 431–452. [https://doi-](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1080/13546783.2013.820220)
2138 [org.ezproxy.lib.bbk.ac.uk/10.1080/13546783.2013.820220](https://doi-org.ezproxy.lib.bbk.ac.uk/10.1080/13546783.2013.820220)
- 2139 Tsai, T., & Gill, J. (2013). Interactions in generalized linear models: Theoretical issues and an
2140 application to personal vote-earning attributes. *Social Sciences*, 2, 91–113
- 2141 MTurk (2017).
2142 <https://docs.aws.amazon.com/AWSMechTurk/latest/AWSMturkAPI/Welcome.html>
- 2143 Vehtari A, Gelman A, Gabry J (2017). Practical Bayesian model evaluation using leave-one-out
2144 cross-validation and WAIC. *Statistics and Computing*, 27, 1413-1432. doi:
2145 10.1007/s11222-016-9696-4 (URL: <http://doi.org/10.1007/s11222-016-9696-4>).
- 2146 Vehtari, A., Gabry, J., Yao, Y., Gelman, A. (2019). loo: Efficient leave-one-out cross-validation
2147 and WAIC for Bayesian models. R package version 2.1.0, <URL: [https://CRAN.R-](https://CRAN.R-project.org/package=loo)
2148 [project.org/package=loo](https://CRAN.R-project.org/package=loo)>.
- 2149 Vineberg, S. (2011). Dutch book arguments. In E.N.Zalta (Ed.), *The Stanford Encyclopedia of*
2150 *Philosophy*, Summer 2011 Edition., available at
2151 <http://plato.stanford.edu/archives/sum2011/entries/dutch-book/>(accessedJan-
2152 uary4,2014).
- 2153 Vul, E., Goodman, N., Griffiths, T. L., & Tenenbaum, J. B. (2014). One and done? Optimal
2154 decisions from very few samples. *Cognitive Science*, 38, 599-637.
- 2155 Ward, W. C., & Jenkins, H. M. (1965). The display of information and the judgement of
2156 contingency. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, 19,
2157 231–241. <https://doi-org.ezproxy.lib.bbk.ac.uk/10.1037/h0082908>

- 2158 Wason, P. C. (1965). The contexts of plausible denial. *Journal of Verbal Learning & Verbal*
2159 *Behavior*, 4, 7-11.
- 2160 Wulff, D. U., Mergenthaler-Canseco, M., & Hertwig, R. (2018). A meta-analytic review of two
2161 modes of learning and the description-experience gap. *Psychological Bulletin*, 144, 140–
2162 176. <https://doi-org.ezproxy.lib.bbk.ac.uk/10.1037/bul0000115>.
- 2163 Yama, H. (2001). Matching versus optimal data selection in the Wason selection task. *Thinking*
2164 *and Reasoning*, 7, 295–311.
- 2165 Yao, Y., Vehtari, A., Simpson, D., & Gelman, A. (2018). Using stacking to average Bayesian
2166 predictive distributions (with Discussion). *Bayesian Analysis*, 13, 917--1007.
2167 doi:10.1214/17-BA1091. <https://projecteuclid.org/euclid.ba/1516093227>

Appendices

2168

Appendix A1

2170 Table A1 shows the best fitting parameter values for the implicit negations data from the
 2171 studies cited in the section *Modelling the default prior*. We used the **DEoptim** function in R
 2172 (Ardia, Mullen, Peterson, & Ulrich, 2016) to find the globally optimal cell values of the JPD
 2173 providing the best fits to the overall frequency of inference endorsements in these studies.

2174

2175 Table A1

2176 *The best-fit parameter value for the four rules in the implicit negations paradigm task.*

| | <i>If p_1 then q_1</i> | | | | <i>If p_1 then $\neg q_1$</i> | | | |
|-------|---|-------|-------|-------|--|-------|-------|-------|
| | 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 |
| | <i>If $\neg p_1$ then q_1</i> | | | | <i>If $\neg p_1$ then $\neg q_1$</i> | | | |
| | q_1 | q_2 | q_3 | Total | q_1 | q_2 | q_3 | Total |
| 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 |

2177

2178

2179

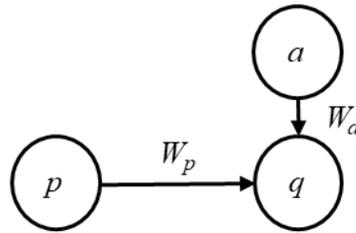
2180

2181

Supplementary Online Material**Causal Bayes nets.**

2183 We have argued that people mentally represent conditionals in a similar way to causal
 2184 Bayes nets (Ali, Chater, & Oaksford, 2011; Ali, Schlottman, Shaw, Chater, & Oaksford, 2010;
 2185 Chater & Oaksford, 2006; Oaksford & Chater, 2010b, 2013, 2016, 2017). Figure S1 shows how
 2186 we can implement the JPD \Pr_1 in Table 1 in a Causal Bayes net where the weights on the directed
 2187 links correspond to causal powers, W_p (Cheng, 1997). In this network *travelling to Manchester* is
 2188 treated as the cause of *Johnny taking the train*, although there may be alternative causes, a , of
 2189 him travelling by train.

2190 Figure S1

2191 *Causal Bayes Net implementing the JPD \Pr_1 in Table 1 interpreted causally*

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

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

2192

2193 In this causal Bayes net, the cause (p) and its alternative (a) are combined using the
 2194 noisy-OR integration rule (Pearl, 1988):

$$2195 \Pr(q = 1|p = 1) = 1 - (1 - W_a)(1 - W_p)^{ind(p)} \quad (\text{Eq. S1})$$

2196 Where $ind(p) = 1$ when the cause is present ($p = 1$) and $ind(p) = 0$ when the cause is absent ($p =$
 2197 0).