Psychologica
2003, **32**, 217-242

Modelling probabilistic effects in conditional inference: Validating search or conditional probability?

Mike Oaksford[*] & Nick Chater[**]

RESUMO

Recentemente, W. Schroyens e W. Schaeken (2003) propuseram uma extensão da teoria dos modelos mentais denominada modelo de procura de validação. Aqueles autores argumentaram que este modelo explica melhor os dados obtidos em tarefas de inferência condicional do que o modelo da probabilidade condicional de M. Oaksford, N. Chater e J. Larkin (2000). Argumentaram, igualmente, que o modelo de procura de validação pode explicar os efeitos probabilísticos na inferência condicional. No presente artigo é avaliada a capacidade destes dois modelos para se ajustarem aos dados obtidos nas experiências 1 e 2 de M. Oaksford et al (2000). Estas são as duas únicas experiências que manipulam explicitamente os parâmetros dos dois modelos. Mostra-se que o modelo da probabilidade condicional se ajusta melhor aos dados do que o modelo de procura de validação. Também são discutidas as consequências da adopção do modelo de procura de validação para a racionalidade humana. Conclui-se que a teoria das probabilidades fornece uma abordagem mais promissora para o raciocínio defectível do dia a dia.

PALAVRAS-CHAVE: Inferência condicional; Modelo de procura de validação; Modelo da probabilidade condicional; Teoria das probabilidades.

The conditional, *if p then q*, is central to any account of human reasoning. Conditionals allow us to describe causal relations, to express rules, hypotheses, and contingent relationships. Once formalised they provide the foundation of logical systems that supply the criteria by which we judge each other's reasoning. However, there is continuing disagreement in the psychology of reasoning on how they are best to be treated. The two main theoretical approaches, mental logic (e.g., Rips, 1994; Braine & O'Brien, 1998) and mental models (e.g., Johnson-Laird & Byrne, 1991), have adopted *material implication* as their computational level theory of the conditional. That is, the conditional is treated as truth-functional: a conditional *if p then q*

[*] Please address all correspondence to: Mike Oaksford, School of Psychology, Cardiff University, PO Box 901, Cardiff, CF10 3YG, UK

Running Head: Modelling Conditional Inference. E-mail: oaksford@cardiff.ac.uk. Tel: + 44 29 2087 4690 (UK (029) 2087 4690)

[**] Department of Psychology, University of Warwick, UK

is true if and only if p is false or q is true. Experimental work has concentrated on the indicative or "straight" (Bennett, 1995) conditional, e.g., if Oswald didn't shoot Kennedy, then someone else did, or, if you turn the key the car starts. In the philosophy of language and logic, "…the majority view [is] that straight conditionals are a matter of subjective conditional probabilities" (Bennett, 1995, p. 332). One consequence of this view is that people, "do not use "if" to express propositions, evaluable in terms of truth" (Edgington, 1995, p. 280). And if conditionals do not have truth conditions then they would not appear to be truth functional, as many psychological theories assume (although this view has its detractors [Lewis, 1976]). Thus, there is a tension in the study of conditionals between normative theory, which treats indicative conditionals probabilistically, and psychological theory, which treats indicative conditionals truth functionally.

However, in line with current normative theories, Oaksford, Chater and Larkin (2000) recently proposed a model of human conditional reasoning based on subjective conditional probability as an extension of their probabilistic treatment of conditional hypotheses in data selection tasks (Oaksford & Chater, 1994, 1996, 1998, in press-a; Oaksford & Wakefield, in press). This model has been shown to provide as good fits to the data on the standard conditional inference task as mental models theory (Oaksford & Chater, in press-b). This is important because the conditional probability model makes no processing assumptions and so provides as a good a fit using only the resources of normative probability theory. This is in contrast to mental models theory, which has to invoke rationally unjustified processing assumptions in addition to standard logic to provide comparable fits to the data. The conditional probability model has also been extended to account for other effects in conditional reasoning (Oaksford & Chater, in press-c), including suppression effects (Byrne, 1989) and order effects (e.g., Evans, Handley & Buck, 1998).

Recent criticisms of the conditional probability model (Schroyens & Schaeken, 2003; for a reply, see Oaksford & Chater, 2003) have argued that probabilistic effects in conditional reasoning might be captured in mental models theory by postulating the existence of a *validating search* process. This process operates after the standard mental models processes have generated a possible conclusion. Participants then search long term memory for a possible counterexample. The probability that a counterexample is available will determine the probability with which an inference is drawn. Schroyens, Schaeken, Fias, and d'Ydewalle (2000) suggest that set-size should affect the availability of counterexamples, the larger the set the more available the counterexample. These authors explored how the number of possibilities may influence set-size. For example, for the rule *if A then 2*, there is only one possibility in which the antecedent and the consequent are both true, i.e., an A, 2 case. However, assuming only the numerals 0 – 9 are in use, there are 9 possible counterexample, where the antecedent is true but the consequent is false, i.e., A paired with any numeral other than 2. Frequency should also affect availability. It may just happen that A, 2 cases simply occur far more frequently than these other possibilities. Frequencies of counterexamples were explicitly manipulated by Oaksford et al (2000). Consequently, in this paper we will explore how well the conditional probability model and the validating search model fare in modelling Oaksford et al's (2000) data.

We begin by introducing the conditional inference task. We then introduce the conditional probability model. We first show how a novel prediction of the model is confirmed by Schroyens and Schaeken's (in press) recent meta-analysis of the conditional inference task. We then introduce Schroyens and Schaeken's (in press) validating search model. These two models are then fitted to the data from Oaksford et al's (2000) Experiments 1 and 2 and their comparative fits assessed.

The Conditional Inference Task

In the conditional inference task participants are presented with a conditional sentence such as *if a bird is a raven (p), then it is black (q)* (the conditional premise), and various facts relating to the antecedent (*p*) or consequent (*q*) of the sentence (the categorical premise). From these pairs of premises logic dictates that various inferences should be made or withheld. So from this rule and the fact that *Tweety is a raven* you should infer that *Tweety is black* (logically this is called *modus ponens*, "MP"). Logic also says that from this rule and the fact that *Tweety is not black* you should infer that *Tweety is not a raven* (logically this is called *modus tollens*, "MT"). Logically you can not infer anything else. However, people are far more willing to make the MP inference than the MT inference, even though there is no logical reason to do so. Moreover, people are also willing to endorse logical fallacies like inferring that *Tweety is not black* from the fact that *Tweety is not a raven* (this is called the fallacy of *denying the antecedent*, "DA") and that *Tweety is a raven* from the fact that *Tweety is black* (this is called the fallacy of *affirming the consequent*, "AC"). The fallacies are endorsed less often than both logical inferences but logically they should not occur at all.

Oaksford et al's (2000) experiments, which we model in this paper, were designed to confirm the conditional probability model's explanation of Evans' negations paradigm (e.g., Evans, 1977, 1993; Evans, Clibbens and Rood, 1995; Evans & Handley, 1999; Evans, Newstead & Byrne, 1993). Therefore, we now introduce this variant on the task. The negations paradigm involves incorporating negations in the antecedents and consequents of the rules. This creates four different task rules (A: Affirmative; N: Negative): *if p then q* (AA), *if p then not-q* (AN), *if not-p then q* (NA), and *if not-p then not-q* (NN), (Evans & Lynch, 1973). This manipulation means that half the conclusions of any inference, MP, DA, AC, or MT will be affirmative and half of them will be negative, i.e., the conclusion will contain a negation. *Negative conclusion bias* is observed when participants endorse more inferences with a negative conclusion than those with an affirmative conclusion. So, for example, in a meta-analysis of the studies cited above, DA was endorsed by only 45.74% of participants when the conclusion was positive, i.e., for AN and NN (a DA conclusion negates the consequent which for these rules is *not-q*, and *not-not-q* is equivalent to *q*). However, DA was endorsed by 69.83% of participants when the conclusion was negative, i.e., for AA and NA. Oaksford et al (2000) showed how a probabilistic model of conditional inference can explain negative conclsion bias. We now outline this model and show how it explains these data.

The Conditional Probability Model

According to the conditional probability model (Oaksford et al., 2000) people will endorse the different inferences depending on their degree of belief in the conclusion given the premises and other background information. MP is straightforward. The conditional degree of belief in the proposition that, for example, birds (B) fly (F) is $P(x \text{ flys} | x \text{ is a bird})$ ($P(\mathrm{F}x|\mathrm{B}x)$). In an MP inference, a categorical premise that affirms the antecedent, is also presented, e.g., Tweety is a bird, and participants must indicate whether they think the conclusion, Tweety flys, can be drawn. So, participants now know that P(Tweety is a bird) = 1, and they are asked what are the chances that Tweety flies, i.e., what is the probability that Tweety flies given that Tweety is a bird, $P(\mathrm{F}x|\mathrm{B}x)$.That is, someone's degree of belief that they can draw the MP should reflect their degree of belief in the conditional:

$$P(\mathrm{MP}) = P(\mathrm{F}x \mid \mathrm{B}x) \tag{1}$$

For the remaining inferences Oaksford et al. (2000) assumed that people have prior beliefs about, for example, the base rates of flying animals, $P(\mathrm{F}x)$, and birds, $P(\mathrm{B}x)$. So in an AC inference, where P(Tweety flys) = 1, Bayes' theorem can be used to calculate the relevant conditional probability $P(\mathrm{B}x \mid \mathrm{F}x)$, that Tweety is a bird given that she can fly, i.e.,

$$P(\mathrm{AC}) = \frac{P(\mathrm{F}x \mid \mathrm{B}x)P(\mathrm{B}x)}{P(\mathrm{F}x)} \tag{2}$$

Conditional probabilities for P(DA) and P(MT) can be calculated, from the fact that the joint probability that something is not a bird and can not fly, $P(\neg\mathrm{F}x \wedge \neg\mathrm{B}x)$, can also be defined in term of $P(\mathrm{F}x)$, $P(\mathrm{B}x)$ and $P(\mathrm{B}x \mid \mathrm{F}x)$.

$$P(\neg\mathrm{F}x \wedge \neg\mathrm{B}x) = 1 - P(\mathrm{F}x) - P(\mathrm{B}x)(1 - P(\mathrm{F}x \mid \mathrm{B}x)) \tag{3}$$

Therefore, the probability that an animal does not fly given it is not a bird (DA) is:

$$P(\mathrm{DA}) = \frac{1 - P(\mathrm{F}x) - P(\mathrm{B}x)(1 - P(\mathrm{F}x \mid \mathrm{B}x))}{1 - P(\mathrm{B}x)} \tag{4}$$

and the probability that an animal is not a bird given that it does not fly is:

$$P(\mathrm{MT}) = \frac{1 - P(\mathrm{F}x) - P(\mathrm{B}x)(1 - P(\mathrm{F}x \mid \mathrm{B}x))}{1 - P(\mathrm{F}x)} \tag{5}$$

In sum, in Oaksford et al. (2000) people were assumed to combine their degree of the belief in the conditional premise with prior beliefs about base rates to derive their degree of belief that the conclusion is true given the truth of the categorical premise. Expressions for the appropriate conditional probabilities for all the inferences can be derived from $P(\mathrm{F}x)$, $P(\mathrm{B}x)$ and $P(\mathrm{B}x \mid \mathrm{F}x)$. More generally, with respect to a conditional, if p then q, we label these probabilities, $P(p)$, $P(q)$, and $P(q \mid p)$. These are the free parameters of the conditional probability model that we fit to the data on conditional reasoning. Although all the conditionals we are going to model can be regarded as implicitly universally quantified, we continue to use the p, q notation standard in the psychological literature. Before turning to Oaksford et al's (2000) explanation of the negations paradigm we point out a further novel prediction of this model.

Positive Correlations Between Inferences

One consequence of this model is that all the inferences should be positively related. This can be seen by briefly introducing the suppression effect in conditional inference (Byrne, 1989; Cummins, 1995). For example, if someone is told that *birds fly* and that *uninjured birds fly* they are less willing to endorse the conclusion that *Tweety flys* given the premise that *Tweety is a bird* because Tweety might be injured. Thus this type of information, called an *additional antecedent*, can *suppress* MP and MT inferences. According to Oaksford and Chater (in press-c) this information reduces the perceived probability that birds fly or equivalently it raises the probability that an animal does not fly even though it is a bird ($P(not\text{-}q \mid p)$). Similarly, if someone is told that *birds fly* and that *bats fly* they are less willing to endorse the conclusion that *Tweety is a bird* given the premise that *Tweety flys* because Tweety might be a bat. Thus this type of information, called an *alternative antecedent*, can *suppress* AC and DA inferences. According to Oaksford and Chater (in press-c) this information increases the perceived probability that an animal does fly even though it is not a bird ($P(q \mid not\text{-}p)$.

According to other theories of the suppression effect, there is no relationship between additional and alternative antecedents. In these theories (see section on the *Validating Search Model* below) the ease of retrieving these cases from long term memory determines whether they influence people's reasoning. Yet no theorist has proposed that the retrievability of additional and alternative antecedents are related. However, according to the conditional probability model the associated probabilities, $P(not\text{-}q \mid p)$ for additional antecedents and $P(q \mid not\text{-}p)$ for alternative antecedents, are positively related. This can be seen by substituting $P(not\text{-}q \mid p)$ for $1 - P(q \mid p)$, and $1 - P(q \mid not\text{-}p)$ for $P(not\text{-}q \mid not\text{-}p)$ in equation (4) (the probability assessed for DA is $P(not\text{-}q \mid not\text{-}p)$ which is $1 - P(q \mid not\text{-}p)$). This yields:

$$P(q \mid \neg p) = \frac{P(q) - P(p)}{1 - P(p)} + \frac{P(p)}{1 - P(p)} P(\neg q \mid p) \qquad (6)$$

That is, $P(q \mid not\text{-}p)$ and $P(not\text{-}q \mid p)$ are linearly related with $\frac{P(q) - P(p)}{1 - P(p)}$ as the intercept and $\frac{P(p)}{1-P(p)}$ as the slope. Oaksford and Chater (in press-c) showed how this means that increases in $P(not\text{-}q \mid p)$ can also affect DA and AC, i.e., increasing the probability of additional antecedents should decrease endorsement of all inferences. Exactly this result was found by George (1997). This suggests that there should be positive correlations not only between the probabilities of drawing the valid inference MP vs. MT and the fallacies DA vs. AC, but also between the valid inferences and the fallacies. No other theory makes this prediction.

We tested this prediction in the meta-analysis of 65 standard conditional inference tasks presented by Schroyens and Schaeken (2003). The mean proportion of each inference endorsed, with study as the unit of analysis, were MP: .97 (SD = .036); DA: .55 (SD = .17); AC: .63 (SD = .20); and MT: .72 (SD = .14). The variance associated with the MP inference was very low. Consequently, correlations with MP might be unlikely to show up. Nonetheless, MP and MT were positively correlated, $r(63) = .23$, $p < .05$, as were DA and AC, $r(63) = .82$, $p < .0001$. Moreover, significant positive correlations were also observed between MT and AC, $r(63) = .37$, $p < .0025$, and between MT and DA, $r(63) = .51$, $p < .0001$. Close to significant positive correlations were also observed between MP and AC, $r(63) = .20$, $p = .061$, and MP and DA, $r(63) = .17$, $p < .084$. These results seem to confirm the prediction of the conditional probability model that the factors that produce the suppression effect are positively related. Moreover, according to the conditional probability account, these factors are always considered in conditional inference regardless of whether additional or alternative antecedents are explicitly introduced, as in our *birds fly* example and in Byrne (1989). Cummins, Lubarts, Alksnis and Rist (1991) and Cummins (1995) reached similar conclusions when they found similar suppression effects *without* explicitly introducing additional or alternative antecedents. Moreover, we will see later on that recent extensions of the mental models approach (see section on the *Validating Search Model* below) also make this assumption.

Probabilities and Negations

Oaksford et al's explanation of the negations paradigm relies on Oaksford and Stenning's (1992) account in which identifying *contrast sets* is one important function of negations. For example, the interpretation of "Johnny didn't serve *coffee*" (where "coffee" is the focus) is that he served a drink other than coffee. The superordinate category "drinks" provides the universe of discourse and the contrast set is defined by the operation of set difference, i.e., it is the set of drinks Johnny could serve, less coffee. This account is called the "otherness" theory of negation which goes back to Plato (see also, Apostel, 1972; Horn, 1989; Ryle, 1929). The set of drinks less coffee is likely to be much larger than the set of coffee drinks. Consequently Oaksford and Chater (1994) suggested that negated categories are treated as high probability contrast sets (higher at least than their un-negated counterparts). The following equivalences were therefore suggested for the rules in the conditional inference task: if *p*, then *q* $\Leftrightarrow$ LL; if *p*, then *not-q* $\Leftrightarrow$ LH; if *not-p*, then *q* $\Leftrightarrow$ HL; if *not-p*, then *not-q* ⇔ HH (where H = high and L = low and the pair, e.g., HL, is ordered to indicate a high $P(p)$ and low $P(q)$ rule). These equivalences allow polarity biases to be re-interpreted. Negative conclusion bias can be regarded as a preference for high probability conclusions and affirmative premise bias can be regarded as a preference for low probability (categorical) premises. Oaksford et al (2000) showed that high probability conclusions are also associated with higher values of the relevant conditional probabilities. So, for example, for the DA inference on *if p then q*, the probability of *not-q* given *not-p* i.e., $P(not\text{-}q \mid not\text{-}p)$, is higher when the probability of *not-q*, $P(not\text{-}q)$, is high. Consequently, high probability conclusions, lead to high probabilities that an inference should be endorsed.

Oaksford et al (2000) varied probabilities in the conditional inference task instead of negations and observed a high probability conclusion effect directly analogous to negative conclusion bias. It is the data from the three experiments in Oaksford et al (2000) to which we fit the different models of conditional inference. We now introduce the validating search model (Schroyens & Schaeken, 2003; Schroyens, et al, 2000).

The Validating Search Model

Schroyens and Schaeken (2003, see also, Schroyens, et al, 2000; Schroyens, Schaeken, & d'Ydewalle, 2001) proposed a validating search model that they regard as a supplement to Johnson-Laird and Byrne's (1991) mental models theory. In mental models theory people interpret conditionals in terms of the cases that make the rule true. People initially represent the conditional, *if p then q*, in working memory either as 7(a) or as 7(b), where 7(a) is the representation for the standard conditional and 7(b) is that for the bi-conditional. For the bi-conditional, not only is it true that *if p then q* but it is also true that *if q then p*.

$$\text{(a)}\quad \begin{array}{ll} [p] & q \\ \ldots & \end{array} \qquad\qquad \text{(b)}\quad \begin{array}{ll} [p] & [q] \\ \ldots & \end{array} \qquad (7)$$

The ellipsis indicates that there might be other conditions that can "flesh out" this representation and the square brackets indicate that p (or q) is exhausted and can not be paired with anything else. Each line in a mental model is like the lines of a truth table. Given the premise p, the representation in 7(a) indicates that the conclusion q is licensed because 7(a) says that all (given by [ ]) the objects that are p are also q. Consequently, people are happy to draw the MP inference. However, nothing can be inferred given the premise $\neg q$ as it does not match anything in working memory. In contrast, in 7(b) the premise q will also suggest the q can only be paired with p, so people will draw the AC inference. MT can be drawn when 7(a) or 7(b) is "fleshed out" with the remaining combinations that make the rule true:

$$\text{(a)}\quad \begin{array}{ll} [p] & q \\ \neg p & q \\ \neg p & \neg q \end{array} \qquad\qquad \text{(b)}\quad \begin{array}{ll} [p] & [q] \\ \neg p & \neg q \end{array} \qquad (8)$$

From 8(a) or 8(b), it can be seen that $\neg q$ can only be paired with $\neg p$, so the MT inference can be made. Moreover, from 8(b) it can be seen that $\neg p$ can only be paired with $\neg q$, so the DA inference can be made.

Schroyens and Schaeken (2003, see also, Schroyens, et al, 2000; Schroyens, Schaeken, & d"Ydewalle, 2001) propose that after people have constructed a mental model of the conditional rule and perhaps fleshed it out, they then perform a validating search of long term memory for potential counterexamples. So for example, their initial model may suggest that they can make the MP inference to the conclusion that *the car starts*, from the premises, *if you turn the key the car starts*, and *you turn the key*. But rather than just go with this conclusion they then search long term memory for a possible counterexample where the car failed to start even though the key was turned (p, $\neg q$). If they find one then either they don't make the inference or they make it with less confidence. Thus, the validating search process supplements the construction and manipulation of mental models.

Schroyens and Schaeken (2003) have parameterised this account and fitted it to their meta-analysis of conditional inference tasks we mentioned above (see section *Positive Correlations Between Inferences*). They provided the following equations for each inference:

$$P(\mathrm{MP}) = 1 - CE_{TF} \qquad P(\mathrm{DA}) = W_{FF}(1 - CE_{FT}) \qquad (9)$$
$$P(\mathrm{AC}) = 1 - CE_{FT} \qquad P(\mathrm{MT}) = W_{FF}(1 - CE_{TF})$$

where CE_{TF} is the probability of finding a p, $\neg q$ counterexample, i.e., $P(p, \neg q)$, CE_{FT} is the probability of finding a $\neg p$, q counterexample, i.e., $P(\neg p, q)$, and W_{FF} is the probability of finding a $\neg p$, $\neg q$ instance, i.e., $P(\neg p, \neg q)$,. CE_{TF} is the probability of finding an additional antecedent (see section *Positive Correlations Between Inferences*), i.e., the probability of finding a bird that does not fly, e.g., an ostrich or an injured bird. CE_{FT} is the probability of finding an alternative antecedent, i.e., the probability of finding an animal that is not a bird but that does fly, e.g., a bat. Thus, the validating search model may also provide an alternative account of suppression effects.

It is also worth noting that apart from the MT inference, the validating search model is formally equivalent to a processing tree (Batchelder & Riefer, 1999) implementation of mental models theory (Oaksford & Chater, in press-a, in press-b). Such a model results in the following equations for each inference:

$$P(\mathrm{MP}) = 1 - P_E \qquad P(\mathrm{DA}) = P_F\,(1 - P_C) \qquad (10)$$
$$P(\mathrm{AC}) = 1 - P_C \qquad P(\mathrm{MT}) = P_F$$

Where P_C is the probability of adopting the conditional interpretation (7a) and P_F is the probability of fleshing out (8). P_E is an error parameter, because according to the mental model account the MP inference should always be drawn. However, it would be unreasonable not to allow for random error (Oaksford & Chater, in press-a). As Oaksford and Chater (in press-b) observed, this means that the models in (9) and (10) are formally equivalent, except for MT, under the following translation rules: $CE_{TF} = P_E$, $CE_{FT} = P_C$, $W_{FF} = P_F$. Consequently, although according to each model, people are engaging in very different activities their ability to fit the data must be very similar.

Schroyens and Schaeken (2003) have fitted the validating search model (9) directly to conditional inference data, even though it was initially introduced as a supplement to mental models theory (Schroyens, et al, 2000; Schroyens, Schaeken, & d"Ydewalle, 2001). Consequently, we will also fit this model directly to the conditional inference data from Oaksford et al (2000). There are two other reasons to do this. First, as we have just seen, it would appear that the mental models component may be redundant. Second, the dependent variables in Oaksford et al's (2000) experiments were inference ratings, which can be interpreted as people's degree of belief that the inference goes through. The validating search model can be interpreted in two ways. Either the probabilities are population parameters or they are individual parameters. On the first interpretation, they represent the proportion of people endorsing an inference. According to this interpretation, whenever people consider an inference, if they find a counterexample, they don't endorse it, otherwise they do endorse it. On the second interpretation, the parameters are individuals' subjective probabilities that a counterexample is available. To turn this into a turn/don't turn decision would require specifying a decision criterion. On the second interpretation, the model should make predictions for the inference ratings in Oaksford et al (2000). Moreover, if this interpretation is ruled out then there may be no mental models account of these data. Alternatively, it could be argued that according to the mental models account people should endorse or fail to endorse each inference maximally. However, in for example, Oaksford et al's (2000) Experiment 1, only one participant failed to use any intermediate values and on average 57% of each participants' ratings were *not* at the extreme ends of the scale. In sum, in order to explain these data the parameters of the validating search model must be interpretable as individual parameters. Consequently, this model should be able to fit these data.

We now turn to the model fitting exercise. We fitted each model to the data from Oaksford et al's (2000) Experiments 1 and 2.

Model Comparison

Oaksford et al's (2000) Experiments 1 and 2 involved explicitly manipulating the probabilities that are supposed to be involved in both the conditional probability model and the validating search model. That is, both experiments provided information about $P(p, q)$, $P(p, \neg q)$, $P(\neg p, q)$, and $P(\neg p, \neg q)$, from which participants could infer $P(p)$, $P(q)$, and $P(q \mid p)$. Schroyens et al (2000) originally articulated the validating search model in the context of the negations paradigm. They pointed out that experiments using alphanumeric material should show set size effects determined by the negations. Let us assume that the only numbers in play are the numerals 0 to 9. Take, for example, the rule *if there is an A then there is a 2*. For this rule there is only one possible p, q case, i.e., A,2, there are 1×9 p, $\neg q$ cases, i.e., A paired with any numeral other than 2, there are 25×1 $\neg p$, q cases, i.e., 2 paired with any letter other than A, and there are 25×9 $\neg p$, $\neg q$ cases, i.e., any letter other than A paired with any number other than 2.

There are two points to make. First, if the antecedent were negated, then for example, the case corresponding to the true antecedent and false consequent would be $\neg p$, $\neg q$ rather than p, $\neg q$. Consequently, rather than 9 possible counterexamples there would now be 225. So negations systematically alter the number possibilities that can act as counterexamples to a rule. Schroyens et al (2000) therefore concluded that the set size of cases is related to the ease of coming up with a counterexample and so to the likelihood that an inference is endorsed. This means that probability or set size manipulations should affect people's reasoning performance.

Second, Schroyens et al (2000) define set size with respect to the number of distinct *possibilities* that are defined. However, although there is only one possibility for the *p*, *q* case for the *if there is an A then there is a 2* rule, these may nonetheless be far more frequent than all other possible letter-number pairings. It only makes sense to allow the number of distinct possibilities to determine the relevant probabilities when participants have no further knowledge of the actual frequencies with which these possibilities occur, as in the standard task version. When these frequencies are known, as they are in Oaksford at al's (2000) Experiments 1 and 2, then the actual frequencies of occurrence determine the probabilities of counterexamples. Consequently, Oaksford et al's manipulations should affect people's inferential performance according to both the conditional probability model and the validating search model.

Another feature of Oaksford et al's (2000) experiments is that they provide more data points to constrain the parameters of the models we will investigate. Both the conditional probability model and the validating search model have three parameters although there are only four data points in the standard conditional inference task without negations. However, in Oaksford et al's experiments, which did not include negations, there are 16 data points for the standard inferences (as in the negations paradigm) plus a further 16 data points for the *converse* inferences. The converse inferences are the standard inferences with the opposite conclusion, e.g., *if p then q*, *p*, therefore, *not-q*. This is the converse of MP, labelled MP′. Thus for each rule participants received eight inferences to perform. The conditional probability model predicts that the probability of endorsing the converse inference, e.g., *P*(MP′), is 1 minus the probability of endorsing the standard inference, i.e., 1 - *P*(MP). Logically, the converse inferences should never be endorsed. Consequently, the standard mental models theory would have to predict that the probability of endorsing these inferences is always 0. Supplementing mental models with a validating search procedure means that the converse inferences might be endorsed because the search for counterexamples raises the possibility of the opposite conclusion. Consequently, we let the validating search model make the same prediction for the converse inferences as the conditional probability model.

We now compare how well the conditional probability model and the validating search model can explain the data from Oaksford et al's (2000) Experiments 1 and 2.

Fitting the models to Oaksford et al's Experiment 1

In the experiments in Oaksford et al (2000) the probabilities of the antecedents and consequents of the condtional rules were systematically varied to produce four rules. Each rule corresponded to one of the four rules in the negations paradigm under the contrast class interpretation (see section *Probabilities and Negations*). There was a low probability antecedent and a low probability consequent rule (LL), a low probability antecedent and a high probability consequent rule (LH), a high probability antecedent and a low probability consequent rule (HL), and a high probability antecedent and a high probability consequent rule (HH). In Experiments 1 and 2 the probability manipulation was achieved by telling participants about the distribution of cards being printed by a machine. In Experiment 1 participants were told that the machine is printing coloured shapes onto cards and sorting them into bins depending on the colour or the shape printed on them. The mean probability ratings for this experiment are shown in the Appendix, Table A1. The details can be obtained from Oaksford et al (2000). What we require are the values of the relevant probabilities that were set in these experiments.

Table 1 – The values of the parameters of the validating search and conditional probability models defined by the probability manipulation in Oaksford et al's (2000) Experiment 1.

	Validating Search				Conditional Probability		
Rule	$P(p,q)$	$P(p,\neg q)$	$P(\neg p,q)$	$P(\neg p,\neg q)$	$P(q\|p)$	$P(p)$	$P(q)$
LL	.167	0	.167	.667	1.0	.167	.333
LH	.167	0	.500	.333	1.0	.167	.667
HL	.200	.467	0	.333	.3	.667	.200
HH	.667	0	.167	.167	1	.667	.833

The values of all the probabilities relevant to the validating search model and the conditional probability model in Experiment 1 are shown in Table 1. The only problematic case is the HL rule, which is problematic because it would appear to be definitely false. That is, most cases are *p*, *not-q* counterexamples. Oaksford et al's (2000) probability manipulation was designed to produce an HL rule, which it has achieved (see, *P*(*p*) and *P*(*q*) values in Table 1). Moreover, in this model comparison exercise we are more concerned with the how well the models fit the data, rather than the relationship with the given probabilities. The given probabilities are related to the best-fit values of the probabilities in these experiments rather like the pi-function in prospect theory (Kahneman & Tversky, 1979). Low probabilities are overestimated and high probabilities are underestimated (see also, Hattori, 2002; Oaksford & Wakefield, in press).

Table 2 – The predicted probabilities of drawing each inference defined by the probability manipulation in Oaksford et al's (2000) Experiment 1 (values for the conditional probability model are in parentheses).

	Inferences			
Rule	MP	DA	AC	MT
LL	1.0(1.0)	**.56(.80)**	.83(.50)	**.67(1.0)**
LH	1.0(1.0)	.17(.40)	.50(.25)	**.33(1.0)**
HL	.53(.30)	**.33(1.0)**	**1.0(1.0)**	.18(.42)
HH	1.0(1.0)	.14(.50)	**.83(.80)**	.17(1.0)

Table 2 shows the predicted probability of drawing each inference in the conditional inference task in Oaksford et al's Experiment 1 (predictions for the converse inferences can be calculated as 1 minus these values). As Table 2 shows, the conditional probability model makes quite extreme predictions because in the probability manipulation there were never any exceptions specified. However, in this experiment participants were told that the rules described faults. That is, they did not describe the normal operation of the machine. This manipulation was designed to get participants to consider the possibility of exceptions. So in the actual results Oaksford et al (2000) predicted less extreme results. Table 2 also shows that in general both models predict the high probability conclusion effect. The values that are in bold correspond to conclusions with high probabilities.

We fitted the model to the data by looking for the parameter values that minimised the sum of squared differences between the observed probabilities of inferences endorsed and the predicted probabilities. We fit the models to each participants data individually. The ratings in the experiment ranged from 1 (definitely do not endorse) to 7 (definitely endorse). To fit the model these ratings were converted to the 0 – 1 probability interval by subtracting 1 and dividing by 6. To estimate best fitting parameter values we minimised the sum of squared differences using a steepest descent search implemented in Mathematica's (Wolfram, 1991) *MultiStartMin* function (Loehle, 2000). This function supplements the Newton-Raphson method with a grid search procedure to ensure a global minimum. We report the fits for the conditional probability model first.

The Conditional Probability Model

According to both models, most of the relevant probabilities should vary for each rule. Consequently, for each rule we defined separate $P(p)$ and $P(q)$ parameters. However, the task only involved a single manipulation of $P(q|p)$ indicating that there may be exceptions (except for the HL rule), i.e., participants were told that the task rules described faults. It was therefore felt appropriate to only allow a single value for $P(q|p)$ for all four rules. Consequently there were nine parameters overall and 32 data points per participant. As indexes of goodness-of-fit, we cite R^2, the proportion of variance accounted for, and RMSD, the root mean squared deviation. There are problems with these measures as indices of goodness-of-fit but we reserve discussion of these problems until the *Discussion* section.

When modelling individual participants data, one can expect some participants' complete data set to be just noise. This may be due to inattention to the task, boredom or the desire simply to get the experiment over with as quickly as possible. We therefore examined participants goodness-of-fit measures. Three participants data could be classified as extreme outliers and their data was excluded from further analysis. Exactly the same three participants were classified as outliers when fitting the validating search model. Consequently, these exclusions do not bias the results in anyway. The average R^2 was .76 (SD = .20; Median = .81), i.e., the average proportion of variance account for in the 32 data points per participant was 76%. The average RMSD was .16 (SD = .09; Median = .15). These are not spectacularly good fits and are worse than the fits reported by Oaksford et al (2000) when modelling just the standard inferences rather than both the standard and converse inferences. However, the purpose of this exercise was comparative, to see which of the validating search model and the conditional probability model provides the better fits to these data.

Table 3 – The predicted probabilities of drawing each inference for each rule in Oaksford et al's (2000) Experiment 1 for the conditional probability model calculated from the average best fit parameter values (SDs in parentheses).

	Inference				Parameter		
Rule	MP	DA	AC	MT	$P(q\|p)$	$P(p)$	$P(q)$
LL	.90	.64	.66	.90	.90(.12)	.43(.24)	.59(.21)
LH	.90	.51	.59	.87	.90(.12)	.44(.25)	.67(.28)
HL	.90	.66	.76	.84	.90(.12)	.55(.25)	.65(.18)
HH	.90	.67	.84	.78	.90(.12)	.65(.23)	.70(.24)

The predicted probabilities of drawing each inference for each rule are shown in Table 3. Table 3 also shows the average best-fit parameter values from which these probabilities were calculated. The observed probabilities of drawing each inference are shown in Table A1 in the Appendix.

Validating Search Model

As for the conditional probability model, because there was only a single general manipulation of exceptions, we used only a single value for the probability of a true antecedent and false consequent, i.e., $P(p, \neg q)$ or CE_{TF}. This also had the consequence that both the conditional probability model and the validating search model had exactly the same number of free parameters, i.e., 9. The model was fitted to the data in exactly the same way as for the conditional probability model. The same three participants were classified as outliers as for the conditional probability model. The average R^2 was .71 (SD = .20; Median = .76). The average RMSD was .19 (SD = .09; Median = .16). The average predicted probabilities of drawing each inference for each rule are shown in Table 4. Table 4 also shows the average best-fit parameter values.

Table 4 – The predicted probabilities of drawing each inference for each rule in Oaksford et al's (2000) Experiment 1 for the validating search model calculated from the average best fit parameter values (SDs in parentheses).

	Inference				Parameter		
Rule	MP	DA	AC	MT	$P(p,\neg q)$	$P(\neg p,q)$	$P(\neg p, \neg q)$
LL	.91	.60	.64	.85	.09(.11)	.36(.15)	.93(.11)
LH	.91	.51	.58	.80	.09(.11)	.42(.22)	.88(.19)
HL	.91	.61	.72	.77	.09(.11)	.28(.15)	.85(.17)
HH	.91	.62	.76	.74	.09(.11)	.24(.15)	.81(.21)

To compare the two models we first performed a paired t-test on the RMSD values for each participant for each model. The RMSD values were significantly lower for the conditional probability model, $t(26) = 3.85$, $p < .0005$, indicating that this model provided significantly closer fits to the data than the validating search model. Moreover, the RMSD values were lower for the conditional probability model than for the validating search model for 21 of 27 participants, $p < .005$ (Binomial test). That is, the conditional probability model provided better fits for significantly more participants than the validating search model.

Looking at Table 4 it is clear that the best-fit parameter values for the validating search model violate the axioms of probability theory. In Oaksford et al's (2000) experiments, participants were given information about the parameters of Schroyens and Schaeken's (2003) model in terms of explicit frequencies. This means that $CE_{TF} = P(p, \neg q)$, $CE_{FT} = P(\neg p,q)$ and $W_{FF} = P(\neg p,\neg q)$ (see also, Oaksford & Chater, in press-b, 2003). Now $P(p,q) + P(p,\neg q) + P(\neg p,q) + P(\neg p,\neg q) = 1$. However, the means of the best-fit parameter values in Table 4 all sum to greater than one for each rule. This means that for each rule $P(p,q) < 0$, which violates the axioms of probability theory. Such *subadditivity* in people's probability judgements is well documented (see, for example, Tversky & Koehler, 1994). However, in this case it seems due not to any irrationality on the part of the participants but due to Schroyens and Schaeken's (2003) particular theoretical formulation, i.e., their parameter values were not constrained to conform to probability theory. Consequently, the validating search model only provides the level of fit it does because it attributes participants in this task with inconsistent beliefs about the availability of the various cases in Oaksford et al's (2000) experimental set up.

If we refit the validating search model to the data under the constraint that the parameters must sum to less than or equal to 1 for each rule, then the comparative model fit is even worse. The RMSD values were significantly lower for the conditional probability model, $t(26) = 7.25$, $p < .0001$ and they were lower for the conditional probability model than for the validating search model for all 27 participants. As Oaksford and Chater (2003) argued, the conditional probability model only relies on the resources of probability theory. That is, the equations of the model are constrained by the rules of probability theory and so the conditional probability model provides a normative computational level theory of the conditional inference task. The validating search model, on the other hand, relies on processing assumptions concerning the availability of counterexamples. In Schroyens et al (2000) it was possible to argue that the probability of accessing counter-examples was not constrained by the frequency of the various truth table cases. However, in Oaksford et al's (2000) experiments this was exactly the information they were given and so accessibility should have been rationally constrained by frequency of occurrence, which in these experiments had to respect the axioms of probability theory.

In sum, for Oaksford et al's (2000) Experiment 1, the conditional probability model provided better fits to the data than the validating search model. In the rest of this main modelling section, we compare the conditional probability model and the validating search model for the two conditions in Oaksford et al's (2000) Experiment 2. In this experiment, the relevant frequencies of the truth table cases were again explicitly manipulated.

Fitting the Models to Oaksford et al's Experiment 2

In Oaksford et al's Experiment 2 participants were told that the machine was printing letters and number on either side of some cards. Again the details can be obtained from Oaksford et al (2000). The mean probability ratings for this experiment are shown in the Appendix, Table A2.

Table 5 – The values of the parameters of the validating search and conditional probability models defined by the probability manipulation in Oaksford et al's (2000) Experiment 2.

	Validating Search				Conditional Probability		
Rule	$P(p,q)$	$P(p,\neg q)$	$P(\neg p,q)$	$P(\neg p,\neg q)$	$P(q\vert p)$	$P(p)$	$P(q)$
LL	.01	0	.01	.98	1.0	.01	.02
LH	.01	0	.98	.01	1.0	.01	.99
HL	.99	0	.001	.009	1.0	.99	.991
HH	.99	0	.009	.001	1.0	.99	.999

The values of all the probabilities relevant to the validating search model and the conditional probability model in Experiment 2 are shown in Table 5. In this experiment there was no attempt to introduce an HL rule, because it was shown in Experiment 1 that people seemed to regard $P(q)$ as higher than $P(p)$ for this, as for the other rules. Oaksford et al (2000) therefore introduced a manipulation predicted by the conditional probability model to lead to the high probability conclusion effect for DA and AC. This can be seen in Table 5 where the small change in $P(\neg p,q)$ and $P(\neg p, \neg q)$ leads to the difference in the predicted probability of endorsing these two inferences between the HL and HH rules. These predicted differences can be seen in Table 6, which shows the predicted probability of drawing each inference in Oaksford et al's Experiment 2. Again, predictions for the converse inferences can be calculated as 1 minus these values.

Table 6 – The predicted probabilities of drawing each inference defined by the probability manipulation in Oaksford et al's (2000) Experiment 1 (values for the conditional probability model are in parentheses).

	Inferences			
Rule	MP	DA	AC	MT
LL	1.0(1.0)	**.97(.80)**	99(.50)	**.98(1.0)**
LH	1.0(1.0)	.0002(.01)	.02(.01)	**.01(1.0)**
HL	1.0(1.0)	**.009(.90)**	**.999(1.0)**	.009(1.0)
HH	1.0(1.0)	.001(.10)	**.991(.99)**	.001(1.0)

As Table 6 shows, both models makes quite extreme predictions because in the probability manipulation there were never any exceptions specified. Moreover, in this experiment, participants were not told that the rules described faults. Consequently, Oaksford et al (2000) argued that participants might consider $P(q|p)$ to be higher in this experiment. Two conditions were used in this experiment, with all participants performing the task in both conditions. The first condition was called the "explicit" condition, because the inferences were presented in standard form, e.g., *if A then 2, not-2*, therefore *not-A* (MP), with the negations explicitly presented. In an "implicit" condition, this same inference would be presented as *if A then 2, 7*, therefore *K*. Participants were also informed that the materials were binary. This manipulation was predicted to make it easier for participants to make MT inferences. Consequently, Oaksford et al argued that the implicit condition should lead to similar rates of endorsement for MP and MT. Table 6 again shows that in general both models predict the high probability conclusion effect. The values that are in bold correspond to conclusions with high probabilities.

We fitted the model in the same way as for Oaksford et al's Experiment 1. The ratings in this experiment ranged from -5 (definitely do not endorse) to 5 (definitely endorse). To fit the model these ratings were converted to the 0 – 1 probability interval by adding 5 and dividing by 10. As before, we report the fits for the conditional probability model first.

The Conditional Probability Model

For the explicit group, two participants data could be classified as extreme outliers and their data was excluded from further analysis. Exactly the same two participants were classified as outliers when fitting the validating search model. The average R^2 was .71 (SD = .22; Median = .75). The average RMSD was .19 (SD = .09; Median = .19). The predicted probabilities of drawing each inference for each rule are shown in Table 7. Table 7 also shows the average best-fit parameter values from which these probabilities were calculated. The observed probabilities of drawing each inference are shown in Table A2 in the Appendix.

Table 7 – The predicted probabilities of drawing each inference for each rule in Oaksford et al's (2000) Experiment 2, Explicit Condition, for the conditional probability model calculated from the average best fit parameter values (SDs in parentheses).

	Inference				Parameter		
Rule	MP	DA	AC	MT	$P(q\|p)$	$P(p)$	$P(q)$
LL	.94	.78	.66	.97	.94(.12)	.31(.19)	.44(.21)
LH	.94	.42	.46	.93	.94(.12)	.34(.25)	.70(.20)
HL	.94	.65	.74	.91	.94(.12)	.51(.22)	.65(.20)
HH	.94	.53	.69	.88	.94(.12)	.53(.21)	.72(.18)

Table 8 – The predicted probabilities of drawing each inference for each rule in Oaksford et al's (2000) Experiment 2, Implicit Condition, for the conditional probability model calculated from the average best fit parameter values (SDs in parentheses).

	Inference				Parameter		
Rule	MP	DA	AC	MT	$P(q\|p)$	$P(p)$	$P(q)$
LL	.94	.65	.70	.93	.94(.10)	.46(.23)	.62(.21)
LH	.94	.47	.56	.91	.94(.10)	.42(.25)	.71(.24)
HL	.94	.66	.77	.90	.94(.10)	.55(.24)	.67(.21)
HH	.94	.50	.70	.87	.94(.10)	.55(.21)	.74(.20)

For the implicit group, one participants data could be classified as an extreme outlier and their data was excluded from further analysis. Exactly the same participant was classified as an outlier when fitting the validating search model. The average R^2 was .74 (SD = .18; Median = .76). The average RMSD was .19 (SD = .08; Median = .18). The predicted probabilities of drawing each inference for each rule are shown in Table 8. Table 8 also shows the average best-fit parameter values from which these probabilities were calculated. The observed probabilities of drawing each inference are shown in Table A3 in the Appendix.

Validating Search Model

To maintain comparability with the conditional probability model we again used only a single value for the probability of a true antecedent and false consequent, i.e., $P(p, \neg q)$ or CE_{TF}. For the explicit condition, the same two participants were classified as outliers as for the conditional probability model. The average R^2 was .68 (SD = .21; Median = .73). The average RMSD was .21 (SD = .08; Median = .20). The average predicted and observed probabilities of drawing each inference for each rule are shown in Table 9. Table 9 also shows the average best-fit parameter values.

Table 9 – The predicted probabilities of drawing each inference for each rule in Oaksford et al's (2000) Experiment 2, Explicit Condition, for the validating search model calculated from the average best fit parameter values (SDs in parentheses).

	Inference				Parameter		
Rule	MP	DA	AC	MT	$P(p,\neg q)$	$P(\neg p,q)$	$P(\neg p, \neg q)$
LL	.94	.65	.68	.89	.06(.12)	.32(.13)	.95(.08)
LH	.94	.40	.46	.82	.06(.12)	.54(.21)	.87(.21)
HL	.94	.62	.70	.84	.06(.12)	.30(.16)	.89(.16)
HH	.94	.52	.61	.81	.06(.12)	.39(.17)	.86(.16)

Table 10 – The predicted probabilities of drawing each inference for each rule in Oaksford et al's (2000) Experiment 2, Implicit Condition, for the validating search model calculated from the average best fit parameter values (SDs in parentheses).

	Inference				Parameter		
Rule	MP	DA	AC	MT	$P(p,\neg q)$	$P(\neg p,q)$	$P(\neg p, \neg q)$
LL	.94	.61	.67	.86	.06(.09)	.33(.17)	.91(.14)
LH	.94	.48	.55	.83	.06(.09)	.45(.26)	.88(.18)
HL	.94	.71	.77	.86	.06(.09)	.23(.17)	.92(.11)
HH	.94	.54	.63	.81	.06(.09)	.37(.17)	.86(.18)

To compare the two models we first performed a paired *t*-test on the RMSD values for each participant for each model. The RMSD values were significantly lower for the conditional probability model, $t(22) = 3.86$, $p < .0005$, indicating that this model provided significantly closer fits to the data than the validating search model. Moreover, the RMSD values were lower for the conditional probability model than for the validating search model for 19 of 23 participants, $p < .0001$ (Binomial test). That is, for the explicit condition in Oaksford et al's (2000) Experiment 2, the conditional probability model provided better fits for significantly more participants than the validating search model.

When we refit the validating search model to the data under the constraint that the sum of the parameters must be less than or equal to 1 for each rule, the comparative model fit is even worse. The RMSD values were significantly lower for the conditional probability model, $t(22) = 6.81$, $p < .0001$ and they were lower for the conditional probability model than for the validating search model for 22 of 23 participants, $p < .0001$ (Binomial test).

We also fitted the validating search model to the implicit condition in Oaksford et al's (2000) Experiment 2. The same participant was classified as an outlier as for the conditional probability model. The average R^2 was .71 (SD = .18; Median = .74). The average RMSD was .20 (SD = .07; Median = .19). The average predicted and observed probabilities of drawing each inference for each rule are shown in Table 10. Table 10 also shows the average best-fit parameter values.

To compare the two models we first performed a paired *t*-test on the RMSD values for each participant for each model. The RMSD values were significantly lower for the conditional probability model, $t(23) = 4.05$, $p < .0005$, indicating that this model provided significantly closer fits to the data than the validating search model. Moreover, the RMSD values were lower for the conditional probability model than for the validating search model for 20 of 24 participants, $p < .0001$ (Binomial test). That is, for the implicit condition in Oaksford et al's (2000) Experiment 2, the conditional probability model provided better fits for significantly more participants than the validating search model.

When we refit the validating search model to the data under the constraint that the sum of the parameters must be less than or equal to 1 for each rule, the comparative model fit is even worse. The RMSD values were significantly lower for the conditional probability model, $t(23) = 8.13$, $p < .0001$ and they were lower for the conditional probability model than for the validating search model for all 24 participants.

In sum, the results of the model comparison exercise for Oaksford et al's (2000) Experiment 2 replicated the results found for their Experiment 1. Participants' performance in Oaksford et al's experiments seems to be modelled more closely by the conditional probability model rather than the validating search model.

Discussion

In this paper, we have addressed the question of how best to explain the effects of quantitative probabilistic manipulations in the conditional inference task. Both the validating search model and the conditional inference model should be capable of explaining these effects because the procedure used by Oaksford et al (2000) explicitly varied the parameters of both models. However, as we have just seen the conditional probability model provided better overall fits, especially when the parameters of the validating search model were rationally constrained such that the probabilities of the different logical possibilities had to sum to 1. In this section, we discuss some possible problems for the model comparison exercise before discussing these results in the context of other recent findings. We close by considering how these results bear on the rationality debate.

We used two indicators of goodness of fit in this paper, the proportion of variance accounted for and the root mean squared deviation. The former provides an indication of how well the model accounts for the overall trend in the data, the latter provides an indication of the average fit to the exact location of each data point. However, as indexes of goodness of fit, these measures do not penalise models for complexity. Complex models may look like they are fitting the data better although they are not doing a better job at capturing the underlying cognitive processes. This is because complex models are able to fit the noise in the data, i.e., they are "overfitting." Model complexity has two features. First, the more parameters a model has the more complex it is. However, in the model comparison exercise we have just reported both models had exactly the same number of parameters. The second feature of model complexity is functional form (Pitt, Myung, & Zhang, 2002). That is, the form of the equations ((1) to (5) for the conditional probability model and (9) for the validating search model) can influence how prone a model may be to overfitting. There are complex statistical procedures (Pitt et al, 2002) that can be used to establish a more level playing field. However, in the current case, there are some other relevant considerations.

First, in contrast to the validating search model, there is no alternative for the functional form of the conditional probability model. Once it is decided that people may be responding in terms of conditional probability then the equations (1) to (5) are set by normative probability theory. On the other hand, the validating search model is implemented as a processing tree model (Schroyens, et al, 2001). The number and order of processes is in the hands of the modeller. This introduces a degree of freedom not open to the conditional probability model. Although, these choices will be reflected in the number of parameters and a processing tree implementation will set a bound on functional complexity. Second, the probabilities in the validating search model attach to the arcs between processes and indicate how likely a process is to be enacted. The resulting formulae (9) do not reflect anything of the complexity of the enacted process. For example, how are we to quantify the complexity of inferring a counterexample from a mental model of the premises? This is an essential part of the validating search model but this additional level of complexity is not reflected in the equations (9). Third, unless rationally constrained, as when we imposed the constraint that the parameters must sum to 1, the parameters of the validating search model are not transparent to experimental manipulation. It is logically impossible to present a frequency manipulation of the parameters of this model, as in Oaksford et al (2000), that would correspond to the best fitting parameter values. The only alternative would appear to be to regard these parameters as arbitrary processing parameters, which are not open to experimental manipulation. This is the same as the probability that a participant will flesh out their initial mental model, which similarly can not be manipulated experimentally.

In sum, if the conditional probability model has a complexity advantage, the validating search model has many other advantages. First, it can be more readily tailored to fit the data. Second, the full complexity of the model is not captured by the equations that are used to fit the data. Finally, if its parameters are not rationally constrained, then they do not have to be transparent to experimental manipulation. These advantages seem to outweigh any that might accrue to the conditional probability model by virtue of functional form.

The results of this model comparison exercise are consistent with recent findings that people are more sensitive to conditional probability manipulations than to the kind of manipulations predicted to effect conditional reasoning by the validating search model (Evans, Handley, & Over, in press). The validating search model suggests that the probability of endorsing the MP inference is 1 minus the probability of an exception. This is consistent with the material conditional of standard logic in which false antecedent cases are true instances of the conditional. So for example, an animal that is not a bird and flys, e.g., a bat, and an animal that is not a bird and does not fly, e.g., a rat, are true instances of the claim that *birds fly*. So according to this account the MP inference should be endorsed in proportion to the probability of true instances.

Evans et al (in press) tested this prediction using a deceptively simple manipulation. According to the conditional probability model, the only cases relevant to whether MP will be endorsed are where the antecedent is true, e.g., the proportion of birds that fly. Now if you know that there are 100 animals, and 10 are birds and 5 fly and 5 don't fly, then according to the conditional probability model, MP will have a .5 probability of being endorsed. However, according to the validating search model (or material implication) it will have a .95 probability of being endorsed. Now if everything remains the same, but there are now only 20 animals in total, the conditional probability will remain unchanged but according to the validating search model MP will now have a .75 probability of being endorsed. Systematically manipulating these probabilities in this way, Evans et al found that people were sensitive to conditional probability but not to the predictions of material implication.

The results of our model comparison exercise are consistent with these results. They show that a conditional probability interpretation can be extended to the other inferences and provide a better explanation of Oaksford et al's (2000) data than the validating search model which is based on material conditional interpretation. Notice that this interpretation extends beyond MP, because of course AC in the validating search model is just MP on the converse conditional, *if q then p*, e.g., all flying animals are birds.

Although the equations provided by the validating search model for MP and AC are consistent with material implication and material equivalance (bi-conditional), the model overall is not consistent with standard logic. This is for the simple reason that if participants are interpreting the task logically and so they assume that the rule is true then according to standard logic there can not be any counterexamples. Consequently, there can be no point in conducting a validating search. The only alternative is that the standard of reasoning that people are attempting to approximate is not given by standard logic but rather by some other normative theory. This means that the addition of the validating search stage marks a move to a different standard of rationality. In standard mental model theory, logically pristine performance is always possible once all the models are fleshed out. However, in this new theory people are not even trying to reason according to this logical standard. This raises two questions: are people reasoning to any normative standard and does it matter?

Oaksford and Chater (2003) argued that the problem whereby the best fit model parameters sum to greater than 1, suggests that people are not trying to reason according to a probabilistic standard in the validating search model. Our model comparison exercise revealed the same behaviour for Oaksford et al's (2000) data. This suggests that some other rational standard may need to be invoked. Alternatively, it may suggest that psychology does not need to concern itself with normative standards. We think this would be a mistake (Chater, Oaksford, Nakisa, & Redington, in press). Showing that people's behaviour conforms to a normative standard explains *why* people behave as they do. It also explains why in general their behaviour is successful. To successfully guide our actions in the real world we need to avoid contradictions and avoid making bets that we are bound to lose, which are guaranteed by logic and probability theory respectively. Of course logically speaking we might have weaker and more tractable goals, which may map onto some weaker non--standard logic, like avoiding contradictions in context but not across contexts. This would suggest that perhaps the psychological processes in the validating search model might approximate a non-standard logical system. One perhaps that allows *defeasibility*, where conclusions may be defeated or suppressed like the conclusion that Tweety flies is defeated by learning (or retrieving from long term memory) the fact that Tweety is an Ostrich.

We have addressed the prospects for defeasible logical systems, either as standards of rationality or as actual psychological mechanisms, many times over the last decade or so (Chater & Oaksford, 1990; Oaksford & Chater, 1991, 1993, 1995, 1998). We have observed that such systems are required to explain many aspects of human reasoning and in particular might be thought to account for suppression effects (for a recent proposal along these lines see, Van Lambalgen & Stenning, 2002). However, as we have also pointed out several times, most of these systems fail to capture many straightforward everyday inferences. For example, suppose you know that Academics are prone to heart disease but that Runners are not, what do you infer when you encounter an Academic runner? Most of us would unequivocally draw the inference that she is unlikely to get heart disease. However, according to defeasible logical systems all that can be inferred is that either she does or she does not get heart disease, i.e., an uninformative tautology. Notice that the reason we unequivocally make the right inference depends on nothing less than our understanding of way the world works. Thus the move to providing a standard of reasoning not based on standard logic or probability theory, is a profound challenge. Without it, it is impossible to say whether the validating search model is capable of explaining why people's reasoning is generally successful.

In conclusion, there are many important consequences of introducing a validating search stage to mental models theory. It makes explicit, what we have argued many times before (e.g., Oaksford & Chater, 1993, 1995, 1998): mental models theory requires an account of how counterexamples are retrieved from memory and a specification of the defeasible logical system that it approximates. These are non-trivial issues. In our opinion, this is the wrong route to go down. Probability theory provides a better account of the relevant data, as we have shown in this paper and elsewhere (Oaksford & Chater, 2003). Moreover, probability theory is compatible with both neural network and Bayesian network representations (Pearl, 1988, 2000) of world knowledge. Although, we do not pretend that this is an easier road to travel we think that we have again shown that it is the one most likely to lead to the desired destination.

Appendix

Table A1 – The average observed probabilities (SD in parentheses) of drawing each inference for each rule in Oaksford et al's (2000) Experiment 1.

	Standard Inferences				Converse Inferences			
Rule	MP	DA	AC	MT	MP′	DA′	AC′	MT′
LL	.91(.19)	.63(.28)	.57(.32)	.88(.19)	.16(.27)	.35(.30)	.38(.31)	.16(.23)
LH	.94(.12)	.41(.31)	.63(.34)	.85(.28)	.08(.22)	.48(.35)	.42(.30)	.10(.19)
HL	.85(.29)	.59(.29)	.72(.28)	.75(.30)	.15(.27)	.33(.29)	.29(.25)	.21(.25)
HH	.93(.11)	.50(.38)	.83(.20)	.80(.30)	.12(.27)	.44(.31)	.15(.15)	.20(.28)

Table A2 – The average observed probabilities (SD in parentheses) of drawing each inference for each rule in Oaksford et al's (2000) Experiment 2, Explicit Condition.

	Standard Inferences				Converse Inferences			
Rule	MP	DA	AC	MT	MP′	DA′	AC′	MT′
LL	.91(.24)	.73(.29)	.53(.36)	.87(.23)	.07(.22)	.24(.21)	.34(.27)	.09(.22)
LH	.93(.20)	.41(.34)	.40(.34)	.76(.34)	.07(.18)	.55(.32)	.54(.37)	.16(.28)
HL	.97(.06)	.59(.33)	.73(.31)	.78(.34)	.09(.23)	.36(.30)	.28(.24)	.05(.13)
HH	.93(.12)	.44(.37)	.67(.33)	.75(.39)	.02(.05)	.48(.31)	.37(.26)	.06(.13)

Table A3 – The average observed probabilities (SD in parentheses) of drawing each inference for each rule in Oaksford et al's (2000) Experiment 2, Implicit Condition.

	Standard Inferences				Converse Inferences			
Rule	MP	DA	AC	MT	MP′	DA′	AC′	MT′
LL	.93(.19)	.53(.37)	.56(.31)	.91(.24)	.11(.22)	.30(.25)	.24(.25)	.14(.29)
LH	.92(.20)	.45(.40)	.45(.36)	.83(.31)	.07(.21)	.50(.31)	.35(.33)	.14(.30)
HL	.97(.06)	.68(.31)	.75(.29)	.91(.19)	.07(.13)	.24(.24)	.20(.19)	.18(.30)
HH	.97(.05)	.38(.32)	.61(.35)	.89(.22)	.09(.23)	.45(.32)	.25(.26)	.15(.31)

RÉSUMÉ

W. Schroyens et W. Schaeken (2003) ont proposé récemment une extension de la théorie de modèles mentaux, dénommée modèle de recherche de validation. Ils soutiennent que ce modèle offre une compréhension des données obtenues avec la tâche d'inférence conditionnelle supérieure à celle qu'on peut obtenir avec le modèle de la probabilité conditionnelle de M. Oaksford, N. Chater et J. Larkin (2000). Ils soutiennent aussi bien que le modèle de recherche de validation permet de comprendre des effets probabilistiques dans l'inférence conditionnelle. Dans cet article, on compare l'aptitude de chacun de ces modèles pour s'accommoder aux données obtenues par M. Oaksford et al (2000). Les expériences 1 et 2 sont l'objet de cet évaluation, car elles sont le seules qui ont explicitement manipulé les paramètres des deux modèles. On montrent que le modèle de la probabilité conditionnelle s'accommode mieux aux données empiriques que le modèle de recherche de validation. Les conséquences de l'adoption du modèle de recherche de validation pour la rationalité humaine sont aussi discutées. On arrive à conclure que la théorie des probabilités offre une approche plus adéquate pour l'entreprise de comprendre le raisonnement défectible de tous les jours.

MOTS-CLÉ: Inférence conditionnelle; Probabilité conditionnelle; Modèles mentaux; Théorie des probabilités.

ABSTRACT

Recently W. Schroyens and W. Schaeken (2003) proposed an extension to mental models theory called the *validating search model*. They argued that this model explains the data on the conditional inference task better than the conditional probability model of M. Oaksford, N. Chater and J. Larkin (2000). They also argued that the validating search model can explain probabilistic effects in conditional inference. In this paper, the ability of these two models to fit the data from M. Oaksford et al's (2000) Experiments 1 and 2 is assessed. These are the only experiments that explicitly manipulate the parameters of both models. It is shown that the conditional probability model provides better fits to the data than the validating search model. The consequences of adopting the validating search model for human rationality are also discussed. It is concluded that probability theory provides a more promising approach to everyday defeasible reasoning.

KEY-WORDS: Conditional inference; Conditional probability; Mental models; Probability theory.
KEY-WORDS: Conditional inference; Validating search model; Conditional probability model.

BIBLIOGRAFIA

Apostel, L. (1972). The relation between negation in linguistics, logic and psychology. *Logique et Analyse*, *15*, 333-401.

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, *6*, 57-86.

Bennett, J. (1995). Classifying conditionals: The traditional way is right. *Mind*, *104*, 331-354.

Braine, M. D. S., & O'Brien, D. P. (1998). *Mental logic*. London: Lawrence Erlbaum.

Byrne, R. M. J. (1989). Suppressing valid inferences with conditionals. *Cognition*, *31*, 1-21.

Chater, N. & Oaksford, M. (1990). Autonomy, implementation and cognitive architecture: A reply to Fodor and Pylyshyn. *Cognition*, *34*, 93-107.

Chater, N., Oaksford, M., Nakisa, R., & Redington, M. (in press). Fast, frugal and rational: How rational norms explain behaviour. *Organizational Behavior and Huamn Decision Processes*.

Cummins, D. D. (1995). Naïve theories and causal deduction. *Memory & Cognition*, *23*, 646-658.

Cummins, D. D., Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. *Memory & Cognition*, *19*, 274-282.

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

Evans, J. St.B. T. (1977). Linguistic factors in reasoning. *Quarterly Journal of Experimental Psychology, 29*, 297-306.

Evans, J. St.B. T., Clibbens, J., & Rood, B. (1995). Bias in conditional inference: Implications for mental models and mental logic. *Quarterly Journal of Experimental Psychology, 48A*, 644-670.

Evans, J. St.B. T., Handley, S., & Buck, E. (1998). Order effects in conditional reasoning. *British Journal of Psychology, 89*, 383-403.

Evans, J. St.B. T., Handley, S. J., & Over, D. E. (in press). Conditionals and conditional probability. *Journal of Experimental Psychology: Learning, Memory & Cognition.*

Evans, J. St.B. T., Newstead, S. E., & Byrne, R. M. J. (1993). *Human reasoning*. Hillsdale, NJ.: Lawrence Erlbaum Associates.

George, C. (1997). Reasoning from uncertain premises. *Thinking and Reasoning, 3*, 161-190.

Hattori, M. (2002). A quantitative model of optimal data selection in Wason's selection task. *Quarterly Journal of Experimental Psychology, 55A*, 1241-1272.

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

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hove, Sussex: Lawrence Erlbaum.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decisions under risk. *Econometrica. 47*, 263-291.

Lewis, D. (1976). Probabilities of conditionals and conditional probabilities. *Philosophical Review, 85*, 297-315.

Loehle, C. (2000). Global Optimization 4.0 [Computer program]. Naperville, IL: Loehle Enterprises.

Oaksford, M., & Chater, N. (1991). Against logicist cognitive science. *Mind & Language. 6*, 1-38.

Oaksford, M., & Chater, N. (1993). Reasoning theories and bounded rationality. In K. I. Manktelow, & D. E. Over (Eds.), *Rationality*, (pp. 31-60). London: Routledge.

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

Oaksford, M., & Chater, N. (1995). Theories of reasoning and the computational explanation of everyday inference. *Thinking and Reasoning, 1*, 121-152.

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

Oaksford, M., & Chater, N. (1998). *Rationality in an uncertain world: Essays on the cognitive science of human reasoning*. Hove, Sussex: Psychology Press.

Oaksford, M., & Chater, N. (in press-a). Optimal data selection: Revision, review and re-evaluation. *Psychonomic Bulletin and Review.*

Oaksford, M., & Chater, N. (in press-b). Conditional probability and the cognitive science of conditional reasoning. *Mind and Language.*

Oaksford, M., & Chater, N. (in press-c). Probabilities and pragmatics in conditional inference: Suppression and order effects. In D. Hardman & L. Macchi (Eds.), *Thinking: Psychological perspectives on reasoning, judgement and decision making*. London: Wiley.

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

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

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

Oaksford, M.. & Wakefield, M. (in press). Data selection and natural sampling: Probabilities do matter. *Memory and Cognition.*

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

Pearl, J. (2000). *Causality*. Cambridge: Cambridge University Press.

Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review, 109*, 472-491.

Rips, L. J. (1994). *The psychology of proof.* Cambridge, MA: MIT Press.

Ryle, G. (1929). Negation. *Proceedings of the Aristotelian Society, 9 (Supp.)*, 80-86.

Schroyens, W., & Schaeken, W. (2003). A critique of Oaksford, Chater, and Larkin's (2000) conditional probability model of conditional reasoning. *Journal of Experimental Psychology: Learning, Memory & Cognition, 29*, 140-149.

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

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

Tversky, A., & Koehler, D. J. (1994). Support theory: A non-extensional representation of subjective probability. *Psychological Review*, *101*, 547-567.

Van Lambelgen, M., & Stennning, K. (2002). *"Suppression effect" and non-monotonicity*. Unpublished Document, Human Communication Research Centre, University of Edinburgh.

Wolfram, S. (1999). Mathematica 4.0 [Computer program]. Cambridge, UK: Cambridge University Press.