Conditional Probability and the Cognitive Science of Conditional Reasoning

MIKE OAKSFORD AND NICK CHATER

Abstract: This paper addresses the apparent mismatch between the normative and descriptive literatures in the cognitive science of conditional reasoning. Descriptive psychological theories still regard material implication as the normative theory of the conditional. However, over the last 20 years in the philosophy of language and logic the idea that material implication can account for everyday indicative conditionals has been subject to severe criticism. The majority view is now apparently in favour of a subjective conditional probability interpretation. A comparative model fitting exercise is presented that shows that a conditional probability model can explain as much of the data on abstract indicative conditional reasoning tasks as psychological theories that supplement material implication with various rationally unjustified processing assumptions. Consequently, when people are asked to solve laboratory reasoning tasks, they can be seen as simply generalising their everyday probabilistic reasoning strategies to this novel context.

Given the centrality of the conditional, *if p then q*, to any account of human reasoning one might have expected some convergence in their treatment across the range of disciplines involved in the cognitive sciences. The explanatory scheme in cognitive science allows an important role both for normative disciplines, such as logic and the philosophy of language, and for the descriptive disciplines such as the psychology of reasoning. For a cognitive scientist, cognition is computation (Oaksford, 1997; Pylyshyn, 1984). Computational explanation is multi-levelled (e.g., Anderson, 1990; Marr, 1982; Pylyshyn, 1984). The computational (Marr, 1982) or rational (Anderson, 1990) level, specifies a normatively justified and descriptively adequate model of some cognitive phenomenon (Oaksford and Chater, 1996, 1998). This is implemented at the algorithmic or performance level. Normally there are aspects of the data that are not explained at the computational level that must be addressed by the algorithmic level. For example, the computational level will not have much to say about speed of processing, which provides an important source of constraint on cognitive theories. Given such an explanatory scheme one might have expected to find cognitive psychologists constructing algorithmic level theories using the current best guess as to the appropriate normative theory of the conditional. Although a reasonable expectation it has not turned out that way.

Address for correspondence: Mike Oaksford, School of Psychology, Cardiff University, PO Box 901, Cardiff, CF10 3YG, UK.
Email: oaksford@cardiff.ac.uk

Mind & Language, Vol. 18 No. 4 September 2003, pp. 359–379.

In the psychology of reasoning, the two main theoretical approaches, mental logic (e.g., Rips, 1994; Braine and O'Brien, 1998) and mental models (e.g., Johnson-Laird and Byrne, 1991), have adopted *material implication* as their computational level theory of the conditional. That is, the conditional is treated as truth-functional. 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. Note though this view has its detractors (Lewis, 1976). Thus, rather than a happy coincidence of opinion, the psychology of reasoning and normative theories of the conditional seem to be passing each other by.

This may be because in the psychology of reasoning the protective belt (Lakatos, 1970) of assumptions defending the truth functional view have not yet been penetrated. These different domains deal with different data. The philosopher of logic and language deals with intuitively compelling examples, whereas the psychologist deals with experimental results. One way of protecting the truth functional view of the conditional against intuitively compelling counterexamples is Grice's proposal that pragmatic conversational factors may intervene. For example, according to material implication, a sufficient condition for believing that 'if the moon is made of cheese, then Mars is green', is that you believe the antecedent to be false, i.e., you believe that the moon is not made of cheese. But this seems to be insufficient grounds for such a belief. However, perhaps examples like this don't count because although this claim has the truth conditions of the material conditional, its *assertability* conditions are different. It has been suggested that this conditional would only be asserted if someone believed that the conditional probability of Mars being green given that the moon is made of cheese is high (Jackson, 1987). (See Edgington (1995), for further arguments against positions like this.)

A way of protecting the truth functional approach has also emerged in the psychology of reasoning. As we have argued, computational explanation in cognitive science is multi-levelled. Given an appropriate computational level theory, there is a choice of cognitive representations and processes in which to implement that theory at the algorithmic level. So, for example, in psychology experiments people are much less likely to endorse inferences corresponding to *modus tollens* (MT, if p then q, $\neg q$, therefore $\neg p$) than *modus ponens* (MP, if p then q, p, therefore q). Mental logicians (e.g., Rips, 1994) explain this finding by proposing that people do not possess an inference schema for MT, but must generate these inferences by the more complex *reductio ad absurdum* argument form. Mental modellers (Johnson-Laird and Byrne, 1991) on the other hand, who adopt a more semantic approach, argue that people do not represent all the conditions that make a conditional true. So they initially represent the conditional, if *p* then *q*, in working memory as,

(1)
$$\begin{array}{ll} [p] & q \end{array}$$

The ellipsis indicates that there might be other conditions that can 'flesh out' this representation and the square brackets indicate that *p* is exhausted and is not paired with anything else. Each line in a mental model is like the lines of a truth table. Note, however, that the uses of '*p*' and '*q*' in mental models theory are not to be confused with propositional variables. Most indicative conditionals used in psychological experiments are implicitly universally quantified, i.e., they relate to a domain of objects that might satisfy the antecedent or consequent. Consequently, the mental models notation treats these symbols in (1) as arbitrary exemplars of the predicates mentioned in the antecedent (*p*) or consequent (*q*). Somewhat confusingly however, when a conditional is specific, e.g., If Oswald didn't shoot Kennedy someone else did, exactly the same notation is used. Given the premise *p*, the representation in (1) indicates that the conclusion *q* is licensed because (1) 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. MT can only be drawn when this representation is 'fleshed out' with the remaining two combinations that make the rule true:

(2)
$$\begin{array}{ll} [p] & q \\ \neg p & q \\ \neg p & \neg q \end{array}$$

From this representation it can be seen that $\neg q$ can only be paired with $\neg p$, so the MT inference can be made.

These examples show that as far as the two major psychological theories of reasoning are concerned material implication provides an adequate computational level theory. All performance deviations are to be explained in terms of the representations and processes that implement logic in the mind/brain. This seems to be at odds with the normative literature where there would appear to be an emerging consensus that the material implication is an inadequate account of everyday indicative conditionals.

A further reason for this situation might be that in the psychology of reasoning, most experiments have been conducted using abstract material, e.g., if there is an A, then there is a 2. Whereas most of the examples used in the philosophy of language and logic fully exploit intuitions based on peoples prior beliefs. The idea behind using abstract material in psychological experiments was explicitly to rule out the influence of people's prior beliefs, to investigate people's reasoning in the 'raw'. However, the problem with this methodology is that it suggests generalising a theory developed to explain these 'toy' reasoning problems to real everyday inferences. But it seems much more likely that people generalise in the opposite direction (Oaksford and Chater, 1993). That is, when confronted with abstract reasoning tasks they import strategies that have been successful in their normal everyday reasoning (Oaksford and Chater, 1993, 1995, 1998, 2002). This suggests that it is normative theories based on 'subjective conditional probability' that we might expect to see in evidence even in abstract reasoning tasks.

Another expectation is that if material implication does provide an adequate computational level theory, then it should also be descriptively adequate, i.e., it should account for most actual reasoning performance in laboratory tasks. That is, one would not expect the algorithmic level theory to have to do too much work in explaining people's reasoning performance. On the other hand, if the algorithmic level has a lot of explanatory work to do, then it must be questionable whether people are really reasoning at all. This is because there is little justification for why these processes are being used other than that they account for the data (Anderson, 1990; Marr, 1982).

In this paper, we discuss recent work that seems to show that conditional probability provides as good an account as algorithmic theories based on material implication such as mental models theory. If normative probability theory is doing as well, or is even close to doing as well, as standard logic supplemented by processing assumptions to capture the data (see above on mental models), then it should be preferred. This is because it will explain more of the data within a rational framework. We present a model fitting exercise that allows us to quantify just how much explanatory work each theory is doing in terms of how much of the variance in the data each model can explain. We look at the data on abstract conditional reasoning tasks. This is because this is where one would have expected the conditional probability model to fair worst, as *prima facie* people's prior beliefs are irrelevant, and hence it would seem that they should be unable to generate the relevant probabilities. Moreover, this is where we would expect theories like mental models to fare best, as it was data like this that these theories were initially designed to explain.

We concentrate on two conditional reasoning tasks. In the conditional inference task (see, Manktelow, 1999, for an overview), people are presented with a conditional, if p then q, and a categorical premise, p, $\neg p$, q, or $\neg q$, and asked whether they would endorse the conclusion, q, $\neg q$, p, or $\neg p$, respectively. That is, they are asked whether they want to endorse MP, denying the antecedent (DA), affirming the consequent (AC), or MT, respectively. Letters and numbers are usually used as the stimulus materials. So people might be confronted with an MP inference framed as follows, given the rule if A then 2, and the fact that A, can you conclude 2? As we discussed above, participants' inferential behaviour on this task does not conform to material implication.

In the selection task (Wason, 1966, 1968) an experimenter presents participants with four cards. Each card has a number on one side and a letter on the other. Participants are also presented with a rule of the form *if p then q*, e.g., *if there is a vowel on one side* (p), *then there is an even number on the other side* (q). The four cards show an 'A'(p card), a 'K'($\neg p$ card), a '2'(q card) and a '7'($\neg q$ card). Participants have to select those cards that they must turn over to determine whether the rule is true or false. Logically participants should select only the p and $\neg q$ cards. However, only 4% of participants make this response, other responses being far more common (p and q cards (46%); p card only (33%), p, q and $\neg q$ cards (7%), p and $\neg q$ cards (4%) (Johnson-Laird and Wason, 1970)).

1. Conditional Reasoning and Conditional Probability

We have been developing an account of the behaviour observed on conditional reasoning tasks based on subjective conditional probability for several years now (Oaksford and Chater, 1994, 1996, 1998, in press; Oaksford, Chater and Larkin, 2000). Despite the dominance of logical based theories in the psychology of reasoning other researchers have also adopted a similar position (Anderson, 1995; Chan and Chua, 1994, George, 1997; Liu, Lo and Wu, 1996; Stevenson and Over, 1995). We now outline how we have modelled the conditional inference task and Wason's selection task. Our presentation of the models is necessarily brief as our purpose is not to argue for these models here. Rather our goal is to show that they provide at least as good an account of the data as logical theories that must be supplemented with arbitrary processing assumptions. Both models have been fully presented elsewhere. The reader is referred to Oaksford and Chater (1994, in press-a) for the full mathematical treatment of our model of the selection task. With respect to our account of the conditional inference task the reader is referred to Oaksford, Chater and Larkin (2000; Oaksford and Chater, 2003a,b).

1.1 The Conditional Inference Task

According to our conditional probability account (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 $P(\text{Tweety is a bird}) = 1$, 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:

$$P(\mathrm{MP}) = P(\mathrm{F}x|\mathrm{B}x) \qquad (3)$$

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 $P(\text{Tweety flys}) = 1$, Bayes' theorem can be used to calculate the relevant conditional probability $P(\mathrm{B}x|\mathrm{F}x)$, that Tweety is a bird given that she can fly, i.e.,

$$P(\mathrm{AC}) = \frac{P(\mathrm{F}x|\mathrm{B}x)P(\mathrm{B}x)}{P(\mathrm{F}x)} \tag{4}$$

Conditional probabilities for $P(\mathrm{DA})$ and $P(\mathrm{MT})$ 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 $P(\mathrm{B}x|\mathrm{F}x)$.

$$P(\neg\mathrm{F}x \wedge \neg\mathrm{B}x) = 1 - P(\mathrm{F}x) - P(\mathrm{B}x)(1 - P(\mathrm{F}x|\mathrm{B}x)) \tag{5}$$

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

$$P(\mathrm{DA}) = \frac{1 - P(\mathrm{F}x) - P(\mathrm{B}x)(1 - P(\mathrm{F}x|\mathrm{B}x))}{1 - P(\mathrm{B}x)} \tag{6}$$

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

$$P(\mathrm{MT}) = \frac{1 - P(\mathrm{F}x) - P(\mathrm{B}x)(1 - P(\mathrm{F}x|\mathrm{B}x))}{1 - P(\mathrm{F}x)} \tag{7}$$

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. Oaksford *et al.*'s (2000) experiments showed that when $P(\mathrm{F}x)$ and $P(\mathrm{B}x)$ were manipulated the above expressions predicted the probability that people endorsed the conclusions of these inferences. Expressions for the appropriate conditional probabilities for all the inferences can be derived from $P(\mathrm{F}x)$, $P(\mathrm{B}x)$ and $P(\mathrm{B}x|\mathrm{F}x)$. More generally, with respect to a conditional, if p then q, we label these probabilities, $P(p)$, $P(q)$, and $P(q|p)$. 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.

We doubt that people make these actual calculations. However, that is a matter for the algorithmic level. What we can legitimately ask is whether such a subjective conditional probability account provides a better computational level theory of people's reasoning performance on the conditional inference task than logic? We can also ask how such an approach fares in comparison to logic based psychological approaches that supplement logic with algorithmic level processing assumptions.

One reason why psychologists show some scepticism about the conditional probability account is that with abstract material and the invocation to assume the truth of the premises, they regard prior knowledge to be irrelevant. Consequently reasoning should be based on a pure logical interpretation of the conditional. There are a few points to make. First, that people don't then show perfect logical performance suggests that, nonetheless, people do not interpret the rule logically. Second, we have argued that people invariably bring prior knowledge to bear in any passage of everyday reasoning (Oaksford and Chater, 1991, 1993, 1995, 1998, 2001). Moreover, there is recent evidence suggesting, 'that even within the context of abstract reasoning problems, people call on their knowledge and beliefs in satisfying task demands' (Schroyens, Schaeken, Fias and d'Ydewalle, 2000, p. 1729). That is, they bring to bear prior, perhaps implicit, knowledge of the specific abstract domain (frequencies of letters and numbers). More interestingly, we suggest that people may bring to bear prior knowledge analogically, i.e., they employ their prior knowledge of using conditional constructions in general. Third, if the subjective conditional probability interpretation is correct, then the conditional premise is not a proposition that can be evaluated in terms of truth (Edgington, 1995, see above). Thus the expectation that asking participants to assume that the rule is true will lead to a particular truth functional interpretation may be misguided. With respect to the conditionals of everyday life, with which participants will be familiar, making this assumption may make no sense. Thus even with abstract material and the instruction to assume that the rule is true, there is no reason why participants should not treat the conditionals in these tasks as analogous to those they normally experience in their everyday lives.

1.2 Wason's Selection Task

Oaksford and Chater (1994) presented a Bayesian optimal data selection (Federov, 1972) account of Wason's selection task. According to this model participants data selection behaviour is rational. Oaksford and Chater's model was based on information gain (Lindley, 1956) although other optimal data selection accounts are also possible (e.g., Evans and Over, 1996; Klauer, 1999; Nickerson, 1996). Here we briefly present a revised version of the optimal data selection model (Hattori 2002, Oaksford and Chater, in press).

In the information gain model people select evidence (i.e., turn cards) to determine whether q depends on p, i.e., $P(q|p) > P(q)$ or $P(q|p) < P(q)$, or whether p and q are statistically independent, i.e., $P(q|p) = P(q)$. So for example, take the hypothesis that *all ravens are black*. People want to find out whether knowing that a bird is a raven increases or decreases the probability that it is black over and above the base rate of black birds. So our optimal data selection model relies on a conditional probability interpretation of the conditional statements used to describe the hypothesis under test in the selection task. Initially participants are assumed to be maximally uncertain about which possibility is true. That is, a prior probability of .5 is assigned to both the possibility of a dependency (the dependence hypothesis, H_D) and to the possibility of independence (the independence hypothesis, H_I). That is, people are assumed to be indifferent between the possibility that the blackness of birds depends, at least in part, on their being ravens, and the possibility that ravenhood is independent of blackness. Participants are looking for evidence that gives the greatest probability of discriminating between these two possibilities. However, the assumption of initial indifference is not crucial: the model's predictions are largely independent of people's prior degrees of belief about which possibility is true (see, Oaksford and Chater, 1994).

In our original dependence hypothesis, $P(q|p) = 1$ (Oaksford and Chater, 1994), i.e., we assumed that in the dependence hypothesis people assumed that there were no non-black ravens. However in the light of the probabilistic treatment of conditionals and causal relations in the normative literature on the philosophy of language and the philosophy of science we believe this to be unrealistic. The selection task was designed as an analogue of scientific reasoning about testing hypotheses about causal laws (Wason and Johnson-Laird, 1972). The probabilistic concept of causal laws has been adopted in the psychology of causal judgement (e.g., Cheng and Novick, 1992). A positive causal dependency (however weak) is taken to exist between p and q as long as $P(q|p) > P(q)$ and a negative causal dependency (however weak) is taken to exist as long as $P(q|p) < P(q)$. That is, a causal dependency exists between p and q as long as they are not statistically independent. In most presentations, our model is parameterised in terms of $1 - \mathrm{P}(q|p)$ rather than $\mathrm{P}(q|p)$, i.e., in terms of the conditional uncertainty of the dependency (see, Edgington, 1995: p. 285). This means that we always assume some non-zero probability that one will find a non-black raven.

Participants' goal is to select evidence (turn cards) that would be expected to produce the greatest reduction in the uncertainty about which of $\mathrm{H_D}$ or $\mathrm{H_I}$ is true. This involves calculating the posterior probabilities of the hypotheses, $\mathrm{H_D}$ or $\mathrm{H_I}$, being true given some evidence. These probabilities are calculated using Bayes' theorem. This requires information about prior probabilities ($\mathrm{P(H_D)} = \mathrm{P(H_I)} = .5$) and the likelihoods of evidence given a hypothesis, for example, the probability of finding an A when turning the 2 card assuming $\mathrm{H_D}$ ($P(\mathrm{A}|2, \mathrm{H_D})$). Given $P(p)$ and $P(q)$ and, in the dependence hypothesis, $P(q|p)$, all these likelihoods can be calculated for each hypothesis. The reduction in uncertainty that can be expected by turning any of the four cards in the selection task can then be calculated. Oaksford and Chater (1994, in press) formalise uncertainty using Shannon-Wiener information, as in Lindley (1956). These papers should be consulted for the mathematical details of the theory. Assuming that the base rates $P(p)$ and $P(q)$ are small (the 'rarity assumption'), i.e., most birds are neither black nor ravens, it turns out that the p and the q cards are expected to provide the greatest reduction in uncertainty about which hypothesis is true. Consequently, the selection of cards that has been argued to demonstrate human irrationality may actually reflect a rational data selection strategy. Indeed this strategy may be optimal in an environment where most properties are rare. A similar Bayesian account of data selection was suggested by Mackie (1963; see also, Horwich, 1982).

2. Comparing Logic and Probability Theory

In this section we discuss some results of fitting the various models to the experimental data on the conditional inference task and Wason's selection tasks. We compare the conditional probability models introduced above with a logical model (based on material implication and equivalence) (Oaksford and Chater, in press, 2003a) and the mental models model (Oaksford and Chater, in press). The data itself is made up of binary judgements about whether a particular inference is valid or not. Thus in modelling these data the assumption is made that the relative frequency of people endorsing a response option reflects the degree of belief that each individual has in the validity of the inference.

The model fits we discuss used two sources of data. First, Schroyens and Schaeken (2003, see also, Schroyens, Schaeken, and Y'dewalle, 2001) reported a meta-analysis of 65 abstract conditional inference experiments reported in the literature since 1971, involving 2774 participants.[1] Second, Oaksford and Chater (1994) reported a meta-analysis of 34 abstract Wason selection task experiments, involving 845 participants. All of these experiments used the standard abstract format that we used to introduce the tasks. Before reporting the results of a comparative model fitting exercise, we introduce the logical and mental models model and show how each was parameterised and fit to the data.

2.1 The Logical Model

According to all logicist psychological theories people may interpret the conditional in the task rule as either material implication (if p then q) or as material equivalence (if p then q *and* if q then p). This is the difference between *if a bird is a raven then it is black*, which does not imply that *if a bird is black then it is a raven*, and *if it is a triangle it has three sides*, which does imply that *if it has three sides it is a triangle*. If they interpret the rule as material implication, then they should endorse MP and MT only, and if they interpret the rule as material equivalence, then they should endorse all four inferences. The proportion of people adopting each interpretation is a free parameter. However, both DA and AC should only be endorsed when the conditional is interpreted as material equivalence. So the best estimate of the probability that someone adopts the material implication interpretation, P_C, is the mean relative frequency of endorsements of DA and AC. Moreover, $1 - P_C$ is the probability of the interpreting the rule as material equivalence. It would be unfair not to allow errors even though according to either interpretation MP and MT must be endorsed. Of course errors may also occur for DA and AC. However, according to material implication an error involves endorsing DA and AC, whereas according to material equivalence an error involves failing to endorse these inferences. We can therefore assume that the errors for DA and AC cancel. The probability of an error, P_E, can then be calculated as 1 minus the mean relative frequency of endorsements of MP and MT. The least means square fits for this

[1] We thank Walter Schroyens and Walter Schaeken for providing us with these data and also with their detailed fits of both the conditional probability model and their validating search model (see below). It is the results of their model fits to the conditional inference task data for these models that we report later on.

logical model are therefore given by the mean of MP and MT for $1 - P_E$ and the mean of AC and DA for $1 - P_C$. In summary, the equations for the probability of drawing each inference in the conditional inference task are:

$$\text{(8)} \quad \begin{array}{ll} P(\text{MP}) = 1 - P_E & P(\text{DA}) = 1 - P_C \\ P(\text{AC}) = 1 - P_C & P(\text{MT}) = 1 - P_E \end{array}$$

To model the selection task does not involve much extra work. If the task rule, *if p then q* is interpreted as material implication then participants should select the *p* and the *not-q* cards. If it is interpreted as material equivalence then participants should select all four cards. Consequently, the model for the conditional inference task carries over directly given the following translation rules: MP = p, DA = $\neg p$, AC = q, MT = $\neg q$. However, in introducing the selection task we pointed out that the dominant response pattern was to select the *p* and *q* cards. As few as 4% of participants select the logically correct cards. Consequently, the fit to the selection task is so poor that we don't report it.

2.2 The Mental Models Model

Oaksford and Chater (in press) parameterised the mental models account we introduced above for the Wason selection task using a simple processing tree model (Batchelder and Riefer, 1999). In mental models theory only the material implication (see example above) and the material equivalence interpretations are allowed. The initial model for material equivalence is [p] [q], which licenses the selection of the *p* and the *q* card. When fleshed out with $\neg p$ $\neg q$, all four cards should be chosen. So according to this theory there are only four possible response patterns: turn the *p* card only, turn the *p* and *q* cards, turn the *p* and the $\neg q$ cards, and turn all four cards. If participants adopt a material implication interpretation then their initial mental model licenses just turning the *p* card. However, if they flesh out then they should turn both the *p* and the *not-q* cards. If they adopt the material equivalence interpretation, then they should turn the *p* and the *q* cards. However, if they flesh out, they should turn all four cards. There are two choice points in processing this information. The first is whether a conditional or a biconditional interpretation is adopted, the second is whether the representation is fleshed out.

Oaksford and Chater (in press) parameterised this account for the selection task in terms of two probabilities: the probability that people adopt a conditional interpretation (P_C) and the probability that they flesh out their initial representation (P_F). In a processing tree model, P_C and P_F are assumed to be independent (Batchelder and Riefer, 1999). This is certainly the simplest instantiation of the model. Moreover, no one, to our knowledge, has ever proposed that these processes are not independent, although investigating models where the independence assumption is not made might be of interest. Oaksford and Chater (in press) also allowed for the possibility of error for the MP inference/*p*-card because it would be unreasonable to expect any psychological theory to be committed to a

100% response rate. We therefore introduced an error parameter as we did for the logical model. Assuming the translation rules introduced in the last section, the equations for the probability of drawing each inference in the conditional inference task are:

$$(9) \qquad \begin{array}{ll} P(\mathrm{MP}) = 1 - P_E & P(\mathrm{DA}) = P_F(1 - P_C) \\ P(\mathrm{AC}) = 1 - P_C & P(\mathrm{MT}) = P_F \end{array}$$

2.3 Comparing the Models: The Conditional Inference Task

For the conditional inference task, each of the models was fitted to the conditional inference data in the same way. The best-fit parameter values were obtained by minimising the sum of squared differences between the models' predictions and each of the four data points for each study in the meta-analyses described above. This was achieved using the *MultiStartMin* function in *Mathematica* (Loehle, 2000; Wolfram, 1991) which uses Newton's method but with a generalised gradient and includes many random starts to ensure the global minimum is found. To compare model fits, in Figure 1 we present scattergrams of the sum of least squares for each model for each study. So in Figure 1 (panel A), for each study the sum of least squares for the mental models model is plotted against the sum of least squares for the logical model. Studies falling in the region above the diagonal line are those for which the sum of least squares is less for the mental models model than logic, i.e., they are the studies for which the mental models model provides a better fit. For those studies that fall below the diagonal line the logical model provides a better fit. As can be seen from panel A the mental models model provided a better fit for 61 of 65 studies and it was a tie for the remaining 4 ($p < .00001$, Binomial Test). On average the logical model only accounted for 56% (SD = 31%) of the variance in each study whereas the mental models model accounted for 88.4% (SD = .13%), $t(64) = 9.49$, $p < .0001$. That is, the mental models model does provide substantial added value by postulating that people only represent part of the meaning of a truth functional conditional and that these initial representations need to be 'fleshed out'. So this analysis show that these algorithmic processing assumptions are needed to account for a good proportion of the variance in the data.

Figure 1 (Panel B) shows how well the conditional probability model does in comparison to the logical model. The conditional probability model provided a better fit for 54 out 65 studies, $p < .0001$ (Binomial Test), and accounted for 84.5% (SD = .14) of the variance, $t(64) = 6.73$, $p < .0001$.[2] That is, even when error is allowed for in the logical model, the conditional probability model fits the data much better. So the conditional probability interpretation provided a rational explanation of significantly more of the data (29%) than logic. This is important,

[2] These fits for the conditional probability model were first reported by Schroyens and Schaeken (in press).

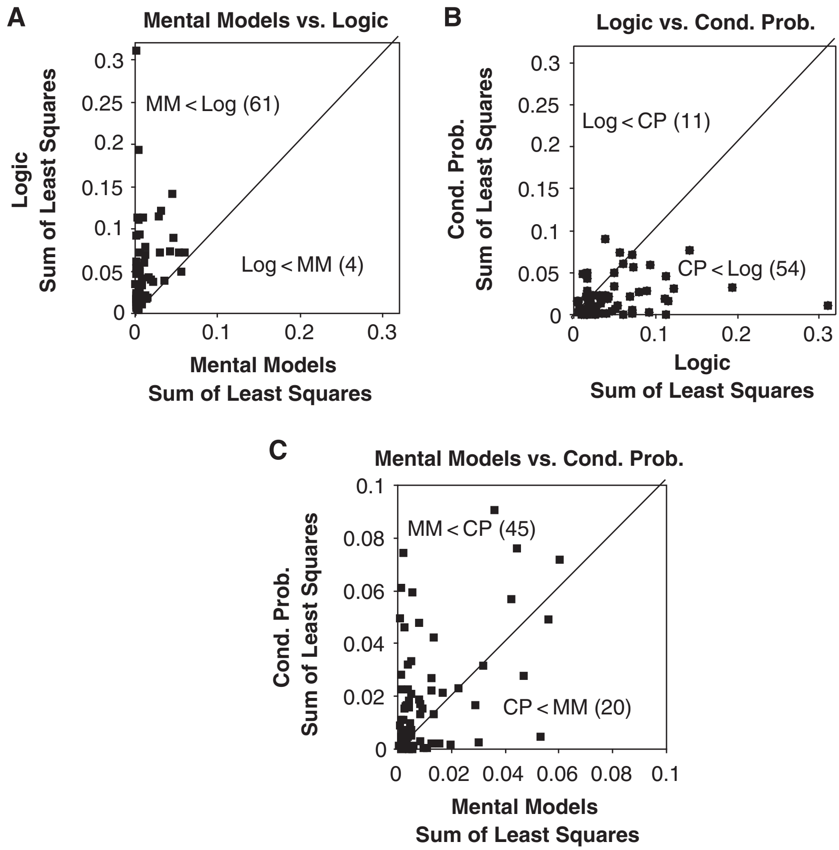

Figure 1 *Scatterplots plotting the sum of least squares for each of the 65 conditional inference studies for the mental models model (MM) vs. the logical model (Panel A), the logical model vs. the conditional probability model (CP) (Panel B), and the mental models model vs. the conditional probability model (Panel C). Note the change of scale for Panel C. In for example, Panel A the region above the diagonal line is where the sum of least squares for MM is less than for the logical model and consequently MM provides a better fit. Below the diagonal line, the logical model provides the better fit.*

but what is more interesting is how the conditional probability model fares in comparison to the mental models model. This comparison is shown in Figure 1 (Panel C; note the change in scale for this panel). The mental models model provided a better fit for 45 out of 65 studies ($p < .003$, Binomial Test). However, there was no significant difference in the mean proportion of variance accounted for ($t(64) = 1.82$, $p = .074$). On average, the improvement in the proportion of variance accounted for by adopting the mental models approach was only 3.9%. That is, the conditional probability model provides almost as good an account of the data on conditional inference, without the need to invoke processing assumptions at the algorithmic level. Thus it seems more natural to argue that people are doing quite well at reasoning probabilistically on this task, rather than that they are trying to reason logically, but are just not doing very well at it because of their limited cognitive mechanisms.

2.4 Comparing the Models: Wason's Selection Task

Oaksford and Chater (in press) originally fit the model to the data by assuming that card choice is independent and minimising the log-likelihood. According to this analysis the mental models model could be rejected, whereas the optimal data selection model could not be rejected. However, for consistency we present the results of these analyses in the same way as we have for the conditional inference paradigm. As we mentioned above, we do not present results for the logical model as it provides such a poor account of people's performance on this task. Figure 2 (same scale as Figure 1, Panel C) shows a similar scattergram to those we presented for the conditional inference task. As can be seen the mental models model fit the data better for 17 studies and the optimal data selection model fit the data better for 17 studies ($p = .5$, Binomial Test). On average the mental models model accounted for 93.1% (SD = 14%) of the variance in each study whereas the optimal data

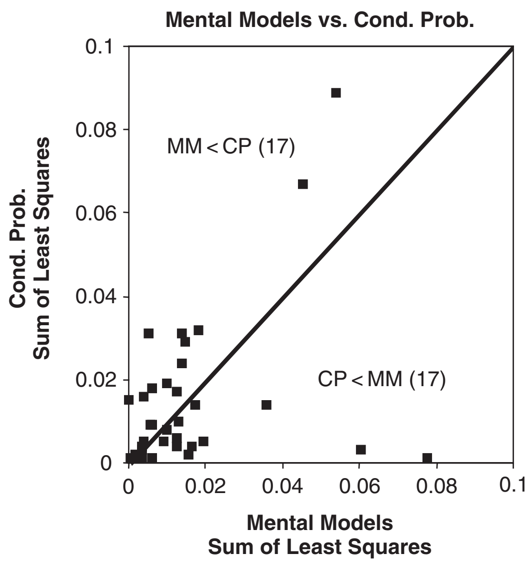

Figure 2 *Scatterplot plotting the sum of least squares for each of the 34 selection task studies for the mental models model vs. the conditional probability model. Note the scale is the same as Figure 1, Panel C. The region above the diagonal line is where the sum of least squares for MM is less than for CP and consequently MM provides a better fit. Below the diagonal line, CP provides the better fit.*

selection model accounted for 95.5% (SD = .06%). This difference was not significant ($t(33) = 1.00$, $p = .324$). As for the conditional inference task, it would seem that the normative conditional probability interpretation can provide as good an account of these data as mental models, without the need for further processing assumptions.

3 Further Observations

We make two further observations. The first involves the values of the best-fit parameter values and the second involves an alternative mental models model.

3.1 Best-fit Parameter Values

The mean values of $P(p)$ (.23, SD = .11) and $P(q)$ (.27, SD = .13) in the selection task were significantly lower than in the conditional inference task ($P(p) = .54$, SD = .15; $P(q) = .73$, SD = .13), $t(97) = 11.14$, $p < .0001$, and $t(97) = 22.25$, $p < .0001$, respectively. The lower values for the selection task are consistent with the rarity assumption (see above). Moreover, Oaksford *et al.* (2000) argued that the probabilities are higher in the inference task because, 'inferences are only relevant when the properties or events to which they apply are more likely than normal to occur'. For example, suppose you know that swans are aggressive. It seems unlikely that this information will be accessed unless someone is in a context where s/he is more likely than normal to encounter swans, i.e., $P(x$ is a swan) is higher than it's default rarity value. While this is true for inference, Wason's selection task engages *inductive* reasoning. The goal is to establish generalities, such as, for example, swans are aggressive, that are generally true across contexts. Across contexts, swans and aggressive things are rare. In sum, although it is rational to take account of rarity in selecting data to test hypotheses, it is usually only in contexts where the antecedent is more likely to be satisfied than normal, that people need to use conditionals embodying world knowledge to draw inferences.

In fitting the optimal data selection model to the selection task data, Oaksford and Chater (in press) fixed $P(q|p)$ at .9. When this parameter was allowed to vary in fitting the conditional probability model to the conditional inference task the mean value of $P(q|p)$ was .89. So achieving good fits to both tasks involves similar assumptions about the levels of conditional dependence that people adopt for these abstract conditional rules. But it has to be conceded that this level of conditional dependence is responsible for the relatively poor fit for the MP inference in the conditional inference task (which is endorsed by 96.9% of participants).

There were also between task differences for the parameters of the mental models model. The mean value of P_C in the conditional inference task (.34, SD = .19) did not differ significantly from the selection task (.37, SD = .11). That is, according to this model roughly the same proportion of participants interpreted the rule as a conditional in both tasks. However, the mean value of P_F was significantly higher in the conditional inference task (.75, SD = .12) than in the selection task (.26, SD = .14), $t(97) = 18.34$, $p < .0001$. That is, according to this model a much higher proportion of participants fleshed out their initial models in the conditional inference task than in the selection task. We could find no theoretical reason to expect such between task differences in the mental models literature.

In fitting the mental models model to the selection task data, Oaksford and Chater (in press) fixed P_E at .11. This was the mean of one minus the probability of selecting the *p* card, which provides the best-fit value. However, in the conditional inference task, the mean value of P_E was .03. So achieving good fits to both tasks for the mental models model also requires the assumption that the level of error is higher in the selection task than in the conditional inference task. One possible interpretation of these effects for the mental models approach is that people find the selection task harder, therefore they process the conditional more superficially (they flesh out less) and they make more errors.

In sum, to explain the data from these two tasks both models must explain significant between task variation in their parameter values. According to the conditional probability model, this is because of the change in the inferential task people confront (Oaksford *et al.*, 2000). The mental models model provides no explanation of this variation other than perhaps attributing it to differences in task difficulty.

3.2 An Alternative Mental Models Model

Recently, Schroyens and Schaeken (2003, see also, Schroyens, Schaeken, Fias, and d'Ydewalle, 2000; Schroyens, Schaeken, and d'Ydewalle, 2001) have proposed an alternative mental models model, although they regard it as a supplement to the mental models account we have addressed in this paper. They 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 will 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 they either 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 (in press) have parameterised this account and fitted it to the same conditional inference data we discussed above. They provided the following equations for each inference:

$$\text{(10)} \quad \begin{array}{ll} P(\text{MP}) = 1 - CE_{TF} & P(\text{DA}) = W_{FF}(1 - CE_{FT}) \\ P(\text{AC}) = 1 - CE_{FT} & P(\text{MT}) = W_{FF}(1 - CE_{TF}) \end{array}$$

where CE_{TF} is the probability of finding a p, $\neg q$ counterexample, CE_{FT} is the probability of finding a $\neg p$, q counterexample, and W_{FF} is the probability of finding a $\neg p$, $\neg q$ instance.

Apart from the MT inference, this model is formally equivalent to the mental models model that Oaksford and Chater (in press) fitted to the selection task, under the following translation rules: $CE_{TF} = P_E$, $CE_{FT} = P_C$, $W_{FF} = P_F$. What this means is that although under each model, people are regarded as engaging in very different activities, their ability to fit the data is very similar. The additional wrinkle for the MT inference does improve the fit but by only 4.3%. This seems to suggest that either constructing and manipulating mental models or the validating search process is explanatorily redundant, either process can adequately explain the data without needing to invoke the other.

Oaksford and Chater (2003a) also point out that the validating search model seems to commit people to probabilistically inconsistent beliefs. It is a reasonable assumption that the probability of recalling a p, $\neg q$ counterexample is not unrelated to the probability that such a counterexample exists, $P(p,\neg q)$. Moreover, as Oaksford and Chater (in 2003a) point out, if Schroyens and Schaeken want to provide a rational basis for their model then this identification is required according to material implication. Similar relationships should hold for the other parameters of their model: $CE_{FT} = P(\neg p,q)$ and $W_{FF} = P(\neg p,\neg q)$. 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 reported in Schroyens and Schaeken (2003) were $CE_{TF} = .048$ $CE_{FT} = .348$, $W_{FF} = .778$. This means that $P(p,q) < 0$, which violates the axioms of probability theory. Such subadditivity in people's probability judgements is well documented (see, for example, Tversky and 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. Their model provies a small (8.2%), but significant, improvement in the proportion of variance accounted for over the conditional probability model. But, as we have just seen, this improvement comes at a cost: people have to be attributed with probabilistically inconsistent beliefs.[3]

However, the theoretical idea on which Schroyens and Schaeken (in press) base their model is related to the intuition that lies at the heart of the conditional probability model. According to the Ramsey thought experiment (Edgington, 1995) in evaluating a conditional, people add the antecedent to their stock of beliefs, make minimal adjustments to establish consistency, and then evaluate the probability of the consequent. So for our everyday conditional about starting cars, one assumes that the key has been turned and then evaluates the probability that the car starts. For such a repeatable event, this latter process presumably involves

[3] Schroyens and Schaeken (2003) might argue that their parameters attach to a processing tree model like the one in which Oaksford and Chater (in press) parameterised the mental models model. Consequently these parameters are supposed to be independent. However, under the interpretation they provide, even if they do attach to the arcs of a processing tree model and so can be independent, they should nonetheless be constrained to conform to probability theory, if they want to provide a rational basis for the validating search process.

implicitly evaluating the proportion of occasions in memory that the car actually starts, which also means accessing the occasions on which it doesn't start.[4] It is this latter process that is the core of Schroyens and Schaeken's (2003) validating search model. Consequently, the intuitions behind the conditional probability model and the validating search model are similar. And Schroyens and Schaeken's (2003) do argue that their model provides a better account of the probabilistic component of human reasoning. The difference between these models is that the equations of the conditional probability model are derived from probability theory. So in fitting the model to the data the only constraints that needed to be placed on the parameter values was that they fell in the 0–1 probability interval. For Schroyens and Schaeken's (2003) model to capture the probabilistic component of human reasoning they would have needed to impose the further constraint that the parameters summed to less than 1.[5]

In sum, Schroyens and Schaeken's (2003) validating search model seems to render the mental models model redundant and is based on similar intuitions as the conditional probability model. In our view the sensible resolution is to join the majority opinion in the normative literature and concede that even in abstract conditional reasoning tasks people interpret the indicative conditional in terms of conditional probability.

4. Conclusion

We have argued that a conditional probability model of the conditional can account as well as other models for the data on abstract conditional reasoning tasks. These are the tasks that processing accounts like mental models theory were designed to explain and where they should show the greatest advantage over the conditional probability approach. However, the mental models model, although providing an advantage over logic alone, failed to reveal a large and significant advantage over normative probability theory. It is important to recall that the algorithmic assumptions made by the mental models model, i.e., partial initial models plus fleshing out, have no rational basis. Thus there is no explanation provided of why these initial models usually result in successful behaviour. Consequently, the conditional probability model should be preferred because it

4 Of course, searching for counter-examples won't be possible when evaluating conditionals about many everyday indicative conditionals, because they won't be about repeatable events. That is, there won't be any specific counter-examples to locate in memory. In which case, other kinds of evidential relationships will have to be established of the type assumed to underlie the assessment of subjective probabilities.

5 If the parameters were allowed to sum to 1, then Schroyens and Schaeken would be committed to a theory where someone believes a conditional rule, if A then 2, even though there are no things that are A and 2. Of course, as mental models is committed to material implication, this is OK, as under this interpretation it makes sense to believe this rule purely on the grounds that there are no As!

provides a rational explanation of a greater proportion (84.5%) of the data (only the 56% explained by the logical model is normatively justified according to mental models theory). We also argued that recent attempts to supplement the mental models approach with a validating search procedure renders the mental models account redundant and attributes people with probabilistically inconsistent beliefs.

Establishing the viability of the conditional probability model for these data is also central to establishing the right pattern of generalisation for psychological theories of conditional reasoning. Normative studies of everyday indicative conditionals seem to show that they are best interpreted in terms of subjective conditional probabilities (Bennet, 1995; Edgington, 1995). We have consistently argued (Oaksford and Chater, 1993, 1995, 1998, 2002) that people generalise their strategies for dealing with everyday inference to the laboratory. The results reported here show that this is a reasonable assumption. However, what of theories like mental models that have been developed to explain laboratory tasks? Given that they are based on a normative theory that has been shown to provide an inadequate account of everyday inference, it seems unlikely that they are going to be able generalise to the inferences that people perform in their everyday lives. Thus trying to make material implication map onto the laboratory data seems misguided when it is already known that the real conditionals of everyday life do not conform to this interpretation. Oaksford and Chater (1993) characterised this strategy as like developing a theory of thinking by modelling the game of monopoly: the principles of such a theory seem unlikely to generalise much beyond that domain.

There are still considerable problems for a cognitive scientific theory of conditional reasoning based on conditional probability. First, there is no explicit algorithmic level account that shows how the theory is implemented in the mind/brain. However, we are working on a neural network implementation that of course may have the advantage of capturing even more of the data. Bayesian network models may also provide a useful formalism in which to implement conditional inference (Pearl, 1988, 2000). Second, although we have concentrated on the abstract data, there is much more psychological data on conditional reasoning. We need to show that the conditional probability interpretation is also consistent with these further results (see, for example, Oaksford and Chater, in 2003b). Importantly, however, there is also data that seem to be consistent only with a conditional probability interpretation. For example, purely probabilistic manipulations in both the selection task (Green and Over, 1997, 2000; Oaksford, Chater and Grainger, 1999; Oaksford, Chater, Grainger and Larkin, 1997) and the conditional inference task (George, 1997, 1999; Oaksford *et al.*, 2000) have been shown to have marked effects on participants inferential behaviour. Third, we need an account of how people can recruit the relevant probabilities from world knowledge (see, Oaksford and Chater, 1998, for discussion of the problems involved).

In conclusion, if the conditional probability model could make no sense of the laboratory tasks, then the cognitive science of conditional reasoning would be in a quandry. It would appear that everyday conditional reasoning obeys one set of principles and reasoning in laboratory tasks obeys a completely different set of principles. Fortunately, we do not have to adopt such an untenable position. It seems that people's everyday strategies based on subjective conditional probability are quite capable of making as much sense of the laboratory reasoning data as alternative processing accounts explicitly designed to explain those results.

School of Psychology,
Cardiff University, UK

Department of Psychology,
University of Warwick, UK

References

Anderson, J.R. 1990: *The adaptive character of thought*. Hillsdale, NJ: LEA.

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

Batchelder, W.H. and Riefer, D.M. 1999: Theoretical and empirical review of multinomial 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. and O'Brien, D.P. 1998: *Mental Logic*. London: Lawrence Erlbaum.

Chan, D. and Chua, F. 1994: Suppression of valid inferences: Syntactic views, mental models, and relative salience. *Cognition*, 53, 217–238.

Cheng, P.W. and Novick, L.R. 1992: Covariation in natural casual induction. *Psychological Review*, 99, 365–382.

Edgington, D. 1995: On conditionals. *Mind*, 104, 235–329.

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

Evans, J. St.B.T. and Over, D.E. 1996: Rationality in the selection task: Epistemic utility versus uncertainty reduction. *Psychological Review*, 103, 356–363.

Fedorov, V.V. 1972: *Theory of Optimal Experiments*. London: Academic Press.

George, C. 1997: Reasoning from uncertain premises. *Thinking and Reasoning*, 3, 161–190.

Green. D.W. and Over, D.E. 1997: Causal inference, contingency tables and the selection task. *Current Psychology of Cognition*, 16, 459–487.

Green. D.W. and Over, D.E. 2000: Decision theoretical effects in testing a causal conditional. *Current Psychology of Cognition*, 19, 51–68.

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

Horwich, P. 1982: *Probability and Evidence*. Cambridge: Cambridge University Press.

Jackson, F. 1987: *Conditionals*. Oxford: Basil Blackwell.

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

Johnson-Laird, P.N. and Wason, P.C. 1970: Insight into a logical relation. *Quarterly Journal of Experimental Psychology*, 22, 49–61.

Klauer, K.C. 1999: On the normative justification for information gain in Wason's selection task. *Psychological Review*, 106, 215–222.

Lindley, D.V. 1956: On a measure of the information provided by an experiment. *Annals of Mathematical Statistics*, 27, 986–1005.

Liu, I., Lo, K., and Wu, J. 1996: A probabilistic interpretation of 'If-then.' *Quarterly Journal of Experimental Psychology*, 49A, 828–844.

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

Mackie, J.L. 1963: The paradox of confirmation. *British Journal for the Philosophy of Science*, 38, 265–277.

Manktelow, K.I. 1999: *Reasoning and Thinking*. Hove, Sussex: Psychology Press.

Marr, D. 1982: *Vision*. San Francisco: W. H. Freeman.

Nickerson, R.S. 1996: Hempel's paradox and Wason's selection task: Logical and psychological puzzles of confirmation. *Thinking and Reasoning*, 2, 1–32.

Oaksford, M. 1997: Thinking and the rational analysis of human reasoning. *The Psychologist*, 10, 257–260.

Oaksford, M. and Chater, N. 1991: Against logicist cognitive science. *Mind and Language*. 6, 1–38.

Oaksford, M. and Chater, N. 1993: Reasoning theories and bounded rationality. In K.I. Manktelow and D.E. Over (eds.), *Rationality* London: Routledge.

Oaksford, M. and Chater, N. 1994: A rational analysis of the selection task as optimal data selection. *Psychological Review*, 101, 608–631.

Oaksford, M. and Chater, N. 1995: Theories of reasoning and the computational explanation of everyday inference. *Thinking and Reasoning*, 1, 121–152.

Oaksford, M. and Chater, N. 1996: Rational explanation of the selection task. *Psychological Review*, 103, 381–391.

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

Oaksford, M. and Chater, N. 2001: The probabilistic approach to human reasoning. *Trends in Cognitive Sciences*, 5, 349–357.

Oaksford, M. and Chater, N. 2002: Commonsense reasoning, logic and human rationality. In R. Elio (ed.),*Common sense, Reasoning and Rationality*. Oxford: Oxford University Press.

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

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

Oaksford, M. and Chater, N. (2003b). Probabilities and pragmatics in conditional inference: Suppression and order effects. In D. Hardman and L. Macchi (eds.), *Thinking: Psychological perspectives on reasoning, judgment and decision making*, 95–122. Chichester, UK: John Wiley and Sons Ltd.

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

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

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

Pearl, J. 1988: *Probabilistic Reasoning in Intelligent Systems*. San Mateo, CA: Morgan Kaufman.

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

Pylyshyn, Z.W. 1984: *Computation and cognition*. Cambridge, MA: MIT Press.

Rips, L.J. 1994: *The Psychology of Proof*. Cambridge, MA: MIT Press.

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

Schroyens, W., Schaeken, W., Fias, W. and 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. and 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.

Schroyens, W., Schaeken, W. and d'Ydewalle, G. 2001b: *A meta-analytic review of conditional reasoning by model and/or rule: Mental models theory revised*. Manuscript submitted for publication (http://www.psy.kuleuven.ac.be/~walters/cognition.pdf).

Stevenson, R.J. and Over, D.E. 1995: Deduction from uncertain premises. *Quarterly Journal of Experimental Psychology*, 48A, 613–643.

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

Wason, P.C. 1966: Reasoning. In B. Foss (ed.),*New Horizons in Psychology*, Harmondsworth, Middlesex: Penguin.

Wason, P.C. 1968: Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, 20, 273–281.

Wason, P.C. and Johnson-Laird, P.N. 1972: *Psychology of Reasoning: Structure and Content*. Cambridge, Mass.: Harvard University Press.

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