

Thinking about conditionals: A study of individual differences

JONATHAN ST. B. T. EVANS, SIMON J. HANDLEY, AND HELEN NEILENS
University of Plymouth, Plymouth, England

AND

DAVID E. OVER
University of Sunderland, Sunderland, England

Recent studies have shown the existence of two qualitatively distinct groups of people based on how they judge the probability of a conditional statement. The present study was designed to test whether these differences are rooted in distinctive means of processing conditional statements and whether they are linked to differences in general intelligence. In the study, each of 120 participants completed three separate cognitive tasks involving the processing of abstract conditional statements—the probability-of-conditionals task, the conditional truth table task, and the conditional inference task—in addition to completing a test of general intelligence (AH4). The results showed a number of predicted effects: People responding with conditional (rather than conjunctive) probabilities on the first task were higher in cognitive ability, showed reasoning patterns more consistent with a suppositional treatment of the conditional, and showed a strongly “defective” truth table pattern. The results include several novel findings and post challenges to contemporary psychological theories of conditionals.

Conditional sentences of the form *if p then q* are everywhere in ordinary language and communication and in technical and scientific discourse. In our view, such statements are of particular importance because *if* is used to initiate the imagination and simulation of possibilities, a process that we term *hypothetical thinking* (Evans, 2007; Evans & Over, 2004). Such imaginary thought is required whenever we entertain a hypothesis, attempt to forecast future events, or imagine the consequences of alternative actions to support our decision making. In view of this, it is not surprising that a huge amount of academic literature has been devoted to the study of *if* in both the philosophy and psychology (see Bennett, 2003; Evans & Over, 2004).

Most contemporary logicians have argued that the *material conditional* cannot be the ordinary conditional of everyday discourse, since this leads to unacceptable paradoxes. The material conditional renders *if p then q* equivalent to *not p or q*. In this case, the statement must be true whenever *p* is false or *q* is true. For example, we would have to believe that the statement “if Al Gore is president then George Bush is president” is true, when it appears quite clearly to be false (in this example, the antecedent is false and the consequent is true). Our own *suppositional* theory of conditionals builds on related psychological work (Braine & O’Brien, 1991; Marcus & Rips, 1979; Oaksford & Chater, 2001; Rips & Marcus, 1977), as well as on studies in philosophical logic (Edgington, 1995, 2003). In compliance with what philoso-

phers call the *Ramsey test* (Ramsey, 1931), we propose that people think about a conditional statement in terms of *p* possibilities, rather than *not-p or q* possibilities. They do this by hypothetically supposing *p* and then running a mental simulation in which they evaluate *q*.

Under the suppositional account, the paradoxical conclusion that the statement “if Al Gore is president then George Bush is president” is true is avoided. We can imagine a world in which Al Gore is president, but in that world George Bush could not also be president, so we would reject the conditional given above. Our work on this theory has led to a theoretical critique of the mental model theory of Johnson-Laird and Byrne (2002), together with the presentation of several empirical findings that we believe favor the suppositional account (Evans & Over, 2004; Evans, Handley, & Over, 2003; Evans, Over, & Handley, 2005; Handley, Evans, & Thompson, 2006; Over, Hadjichristidis, Evans, Handley, & Sloman, 2007). We believe that the model theory is at least partially committed to a material conditional, although this has been denied by some of its supporters (Schroyens & Schaeken, 2004). For the present purposes, however, we can focus on the noncontroversial claim that the model theory (1) does not support the suppositional view of conditionals and (2) differs from our account in the prediction of certain key phenomena. These will be explained below.

In this article, we will confine our interest to *basic* conditionals. Such conditionals concern abstract re-

J. St. B. T. Evans, jevans@plymouth.ac.uk

relationships and are indicative in form, making claims that could apparently be empirically verified, such as “if there is an A on the card then there is a 4 on the card.” Although the most commonly used psychological tasks to study such conditionals are the conditional inference task and the Wason selection task, these do not yield predictions that readily separate the material and suppositional conditional (see Evans & Over, 2004). Hence, we will concentrate our attention initially on the two tasks that have been identified as generating data patterns that specifically support the suppositional conditional and distinguish it both from the material conditional and from the conditional specified by Johnson-Laird and Byrne (2002) in their mental models account (Evans et al., 2005). These are the *truth table task* and the *probability-of-conditionals* task.

Truth table tasks have been studied with basic conditionals for many years (for reviews, see Evans, Newstead, & Byrne, 1993; Evans & Over, 2004). First, there is an evaluation task in which people are presented with all logical cases of a conditional to evaluate, as in the following case:

If the letter is B then the number is 4.
 B4 (TT)
 B7 (TF)
 G4 (FT)
 L8 (FF)

The four logical cases, of which examples are shown above, are known as TT, TF, FT, and FF according to whether the antecedent and consequent components are verified or falsified, respectively, by the case given. Participants have to judge whether each case makes the conditional true or false or is irrelevant to it. If the conditional is material, people should classify TT, FT, and FF as true and TF as false, but this is rarely observed. Much more common is the defective truth table pattern originally proposed by Wason (1966), in which people classify TT as true, TF as false, and both FT and FF as irrelevant (Evans et al., 1993). From the viewpoint of the suppositional conditional, this pattern is to be expected: If the truth of the conditional is determined by the Ramsey test, then not- p cases are indeed irrelevant. Supporters of the mental model theory, however, have strongly denied the psychological reality of the defective truth table. We will consider Johnson-Laird and Byrne’s (2002) explanation of the effect below.

The second task that is claimed to distinguish the suppositional theory from the mental models account is the probability-of-conditionals tasks (Evans et al., 2003; Evans, Ellis, & Newstead, 1996; Girotto & Johnson-Laird, 2004; Hadjichristidis et al., 2001; Oberauer & Wilhelm, 2003). In a typical task, participants are given information about the relative frequency of TT, TF, FT, and FF cases and are asked to judge the probability that the conditional statement is true. The frequency distributions presented allow computation of several different probabilities that could form the basis for the judged probability of the conditional statements. On such tasks, choices are generally

inconsistent with the probability of the material conditional— $P(\text{not-}p \text{ or } q)$ —although Girotto and Johnson-Laird (2004) reported some cases. The most common response is that predicted by the suppositional theory, to give the conditional probability $P(q|p)$. In compliance with the Ramsey test, people will focus only on the p cases and base their belief in the conditional statement on the likelihood of q under that supposition.

In the study of abstract or basic conditionals, there is a strong minority pattern, with clear evidence of qualitative individual differences demonstrated in two independent studies (Evans et al., 2003; Oberauer & Wilhelm, 2003). Although a majority group of participants consistently use conditional probability, $P(q|p)$, a sizable minority give conjunctive probability, $P(p\&q)$, instead. The proportion of conjunctive responses was somewhat higher in a smaller scale study reported by Girotto and Johnson-Laird (2004). The mental model theory can provide an account of the conjunctive pattern in accordance with the principle of implicit models, if people focus only on the initial pq model. However, there seems to be no theoretical basis for model theory to predict the majority conditional probability response (see below).

In this article, we report a single but large-scale experiment in which 120 participants were given three different cognitive tasks involving abstract conditionals, as well as a test of general intelligence. The three tasks were the truth table task, the probability-of-conditionals task, and the conditional inference task (of which more below). The purpose of the experiment was (1) to test whether the qualitative individual differences observed in the probability-of-conditionals tasks would predict corresponding differences in performance on other tasks involving conditional statements and (2) to provide further evidence relevant to distinguishing the suppositional from the mental model theory of conditionals. We note that Oberauer, Geiger, Fischer, and Weidenfeld (2007) have recently reported a closely related study in which performance on probability judgement, inference, and truth table tasks was compared across individuals, albeit with a number of methodological differences. Since this work came to our attention after the present study had been completed and played no part in our framing of the predictions, we will defer consideration of Oberauer et al.’s study to the Discussion section. We will now present our theoretical rationale and the derivation of our predictions.

Derivation of Theoretical Predictions

The conditional probability response—the majority pattern reported by both Evans et al. (2003) and Oberauer and Wilhelm (2003)—is clearly compatible with the suppositional theory of conditionals, for reasons explained above. Neither group of authors was able to reconcile this with the mental model theory. To understand why, consider this statement:

If the candidate is well qualified then she will be appointed.

According to Johnson-Laird and Byrne (2002), this sentence would be represented initially with only one explicit mental model:

candidate well qualified candidate appointed
 . . .

where “. . .” is an implicit model relating to cases in which the candidate is *not* well qualified. The model theory can allow that people forget about this implicit model and focus only the explicitly represented case of pq and, on this basis, can account for the minority conjunctive pattern. Forgetting the implicit model is also the account given by Johnson-Laird and Byrne (2002, p. 657) for the defective truth table pattern, in which participants frequently rate false antecedent cases as irrelevant, a point to which we shall return.

Those who process conditionals more deeply, according to model theory, flesh out the fully explicit set of possibilities consistent with the material conditional—that is,

candidate well qualified	candidate appointed	(TT)
candidate not well qualified	candidate appointed	(FT)
candidate not well qualified	candidate not appointed	(FF)

The model theory assumes a principle of truth in which only true possibilities are explicitly considered. In order to explain the conditional probability response, however, it appears that we have to assume that people (1) disregard the not- p cases, in line with the defective truth table, and (2) consider the likelihood of the TF, as well as TT, cases. The combination of the principle of truth and the explicit rejection of the defective truth table by Johnson-Laird and Byrne (2002) makes this explanation—and the prevalence of judgments based on $P(q|p)$ —clearly inconsistent with the model theory. Moreover, those who flesh out all the models might be expected to give the material conditional probability, which has rarely been reported in the relevant studies.

The suppositional theory, however, has the converse problem: how to explain the conjunctive pattern. Sperber, Cara, and Girotto (1995) have argued that shallow processing of a conditional sentence will lead only to a focus on p and q (TT) cases and that deeper processing is required to understand that p and not- q (TF) cases are not permitted. Evans et al. (2003) offered an explanation of the conditional/conjunctive probability difference along similar lines. The complete Ramsey test required consideration and comparison of the frequency of TT and TF cases, but some participants are shallow processors who do not think beyond the TT case, ending up with the conjunctive probability. In effect, these participants fail to process the conditional in a suppositional manner. We assume that such superficial processing is more likely to occur in those of lower working memory capacity and cognitive ability, especially when reasoning with abstract problem materials. Hence, our first hypothesis is that conditional responders will have higher AH4 intelligence scores than will conjunctive responders.

Individual differences in suppositional reasoning.

The main purpose of this study was to see whether there are two different ways of processing conditional sentences that affect not only the probabilistic truth table task, but also other tasks typically used to measure conditional reasoning. We predict that those participants who process the conditional statement suppositionally on one task will also do so on another. In particular, the same participants who give the conditional probability response to the probability-of-conditionals task should also provide the most defective truth table patterns on the truth table task.

Although irrelevant judgments are common on the truth table task, there are only two factors known to cause them: (1) the presence of false antecedent cases and (2) cases that mismatch the lexical content (see the following section on matching bias). In our suppositional theory, the former trend is attributed exclusively to suppositional processing of the conditional statement (see Evans & Over, 2004). Hence, on the assumption that conjunctive probability responders are not reasoning suppositionally, we can expect them to give irrelevant responses only for Reason 2. Therefore, we must predict lower rates of defective truth tables for this group. The most obvious alternative pattern that they might produce is the truth table for the conjunction p and q . Hence, they might classify the conditional as though it were conjunctive, regarding TF, FT, and FF cases as false. However, it is also possible that they might regard p and q as equivalent and give a biconditional pattern, classifying TT and FF as true and TF and FT as false. Hence, we predict a less defective pattern for conjunctive probability responders but leave open the issue of what alternative form of response will result.

If we can show that one group of participants, *higher in cognitive ability*, provide more suppositional patterns on the two tasks, this will create major problems for the model theory account. It is well known that cognitive ability, reasoning, and working memory capacity are closely correlated (Colom, Rebollo, Palacios, Juan-Espinosa, & Kyllonen, 2004; Kyllonen & Christal, 1990). Johnson-Laird and colleagues have frequently appealed to working memory capacity as an explanation of errors in reasoning, on the grounds that working memory limits the number of mental models that can be apprehended (Johnson-Laird & Byrne, 1991). On this basis, Johnson-Laird and Byrne must predict that participants of high cognitive ability will be the ones most able to flesh out models of the conditional statement. Also, those who forget about the implicit models on one task (judged probability of the conditional) must be expected to do so on another (the truth table task). Thus, whereas we predict that more false antecedent cases will be classified as irrelevant by conditional probability responders and those higher in cognitive ability, model theorists must predict the *opposite*. It is the high-ability people who should be most able to make determinate classifications of such cases and the low-ability people who will fail to flesh them out and, hence, will judge them irrelevant.

In the analysis that follows, we divided the defective truth table into two components. Logically, participants should classify TT cases as true and TF cases as false

whether conditionals are suppositional or material. Hence, we combined these two judgments to compute a logic index. We predicted that this index would be higher for those of higher cognitive ability. The defective truth table, predicted by the suppositional theory, involves classification of FT and FF cases as irrelevant. We also counted the number of such judgments for each participant and scored this as a false antecedent index (FAI). Both indices were computed across the four kinds of conditional statements, hence controlling for matching bias within the negations paradigm (see the next section).

Individual differences in matching bias. In this study, we employ the truth table task with the negations paradigm, in which conditionals are presented in four forms, with the presence of negations varied in either component (see Table 1). This methodology is important for control purposes, as we shall explain, but also adds the potential for new and interesting findings. In the case of the truth table task, there is a very powerful effect of *matching bias* (Evans, 1998), which is revealed by the use of the negations paradigm. Matching cases are ones that feature the lexical content that is explicit in the conditional statement. Consider the sentence

- (1) If the letter is G then the number is not 4.

When evaluating this statement, participants usually correctly classify G4, the TF case, as false. This is also the double-matching case. A case like T7, which mismatches both antecedent and consequent and is logically the FT case, is normally described as irrelevant. Now consider this statement:

- (2) If the letter is not G then the number is 4.

The TF case is now represented as a double mismatch—for example, T7—which most participants describe as irrelevant. They are much more likely to say that G4, the double-matching but FT case, would falsify Statement 2. A full demonstration of how the negations paradigm works for the truth table task is shown in Table 1. Note that any case marked as pq matches in both antecedent and consequent, whereas the $\neg p\neg q$ case is a double mismatch. Since matching bias also affects the perceived relevance of cases on the truth table task, it needs to be controlled in order for us to test our key predictions about individual differences in the defective truth table.

Although matching bias has been investigated in the psychological literature since the 1970s, there have been no reports of systematic individual differences in the strength of the bias. However, there are few large-scale

Table 1
Matching Cases and Truth Table Cases, Illustrating the Negations Paradigm

Rule	TT	TF	FT	FF
If B then 3	B3	B7	N3	G4
If p then q	pq	p¬q	¬pq	¬p¬q
If G then not 6	G9	G6	T4	A6
If p then not q	p¬q	pq	¬p¬q	¬pq
If not R then 5	M5	B6	R5	R9
If not p then q	¬pq	¬p¬q	pq	p¬q
If not E then not 1	K7	D1	E3	E1
If not p then not q	¬p¬q	¬pq	p¬q	pq

Note—The lexical content negates components of the conditional statement implicitly by a mismatching case; for example, a 7 implies not a 3, and so on.

studies of individual differences in conditional reasoning, and those that have been reported (Klaczynski & Daniel, 2005; Newstead, Handley, Harley, Wright, & Farelly, 2004) have not included the truth table task. Hence, in addition to testing for the predicted individual differences in defective truth table patterns, we can also test for differences in matching bias. By analogy to claims that higher ability participants are less vulnerable to belief biases and other biases on cognitive tasks (Stanovich, 1999; Stanovich & West, 2000), we test the hypothesis that higher ability reasoners (and hence, conditional probability responders) will be less susceptible to matching bias.

Individual differences in conditional inferences. In this study, we also included the conditional inference task, which is one commonly used to study reasoning with abstract conditionals. This task was also presented using the negations paradigm, as is common in the literature (see Table 2). There are four conditional inferences, which we illustrate with an affirmative conditional as follows:

Modus Ponens (MP)

If the letter is B then the number is 4.
The letter is B, therefore the number is 4.

Modus Tollens (MT)

If the letter is B then the number is 4.
The number is not 4, therefore the letter is not B.

Denial of the Antecedent (DA)

If the letter is B then the number is 4.
The letter is not B, therefore the number is not 4.

Affirmation of the Consequent (AC)

If the letter is B then the number is 4.
The number is 4, therefore the letter is B.

Table 2
Logical Form of the Four Conditional Inferences for Statements With Use of the Negations Paradigm

Rule	MP		DA		AC		MT	
	Given	Conclude	Given	Conclude	Given	Conclude	Given	Conclude
If p then q	p	q	not-p	not-q	q	p	not-q	not-p
If p then not q	p	not-q	not-p	q	not-q	p	q	not-p
If not p then q	not-p	q	p	not-q	q	not-p	not-q	p
If not p then not q	not-p	not-q	p	q	not-q	not-p	q	p

Note—MP, modus ponens; DA, denial of the antecedent; AC, affirmation of the consequent; MT, modus tollens.

The conditional inference task involves asking participants whether or not each of the four inferences (MP, DA, AC, and MT) logically follows. Both suppositional and mental model theories agree that MP and MT are valid inferences and that DA and AC are fallacies, whose conclusions are not necessitated by their premises. In general, research with the abstract conditional inference task shows that people more readily make MP than MT and frequently endorse the fallacious inferences, especially AC, when negations are controlled for using the negations paradigm (Evans, Clibbens, & Rood, 1995; Schroyens, Schaeken, d'Ydewalle, 2001). As with the truth table task, negations can affect reasoning on this task—in particular, due to a double negation effect. Both MT and DA inferences tend to be inhibited when a negation needs to be denied to draw their conclusion. For example, the following DA inference—

If the letter is B then the number is not 4.

The letter is not B, therefore the number is 4.

—is made much less frequently than are the inferences in the examples given above for the affirmative conditional.

We expected that shallow processors of the conditional statement (conjunctive probability responders, lower cognitive ability) would be more inclined to make AC inferences. We based this prediction on the fact that in making the conjunctive probability response, they appear to focus their attention on the *pq* or TT case. This should lead to highly frequent MP and AC inferences. However, MP inferences would also be high for those who processed the conditional more deeply, so we expected the difference to show only on AC. We considered earlier the possibility that conjunctive probability responders might reason in a biconditional manner, in which case we might expect them and lower ability participants to make more DA inferences also. This would be consistent with the findings of Newstead et al. (2004) and De Neys et al. (2005), who found that participants of lower working memory span made more DA, as well as AC, inferences.

Newstead et al. (2004) also found that whereas higher ability participants made more MP inferences, they curiously made significantly fewer MT inferences, which are also logically valid. Newstead et al. suggested that there may be an easy pragmatic route to MT used by low-ability participants that is not available to those of higher ability who are engaged in genuine deductive reasoning. In this case, the results are inconsistent with those of De Neys et al. (2005), who found more MT inferences made by those of higher cognitive ability. Our study also provides an opportunity to examine this issue.

Summary. In summary, we predicted that conditional probability responders, as compared with conjunctive probability responders, would be of higher cognitive ability, less prone to AC on the conditional inferences task, and more likely to show a defective truth table pattern. In addition to these specific predictions, our study provides the opportunity to replicate individual differences in conditional inference reported by Newstead et al. (2004) and to test whether there are individual differences in suscep-

tibility to matching bias and double negation bias that are related to cognitive ability.

METHOD

Design

A single group of participants was administered the AH4 test of general intelligence, followed by three separate tasks in a fixed order of testing: the conditional truth table task, the conditional inference task, and the probability-of-conditionals task. Each task involved manipulation of several independent variables, as indicated below. All the tasks were in pencil-and-paper format, and the participants were tested in small groups of up to a maximum of 8 in a single session of approximately 1-h duration.

Participants

One hundred twenty participants were recruited from the population of undergraduate students at the University of Plymouth in exchange for cash payment (39 male, mean age = 23.46 years, $SD = 7.15$; 81 female, mean age = 22.38 years, $SD = 7.12$).

AH4 General Test of Intelligence

The AH4-Group Test of General Intelligence was administered (Heim, 1970), which is designed for use with a cross-section of the adult population. Test-retest reliability has been reported at .919, with retesting after 1 month. The test consists of two parts, each containing 65 items. Part 1 comprises both verbal and numerical problems; Part 2 comprises problems in diagrammatic form. Correlations between scores on Part 1 and scores on Part 2 have been reported to range from .60 to .81 (Heim, 1970). Each section is timed and must be completed in 10 min. Note that the AH4 does not yield an IQ score and is not normalized around a population mean of 100. The observed mean for our participants (95.4, $SD = 13.62$) is close to the available norms for university students.

Truth Table Task

The participants were presented with a booklet containing problems involving eight conditional statements, two each of the general form *if p then q*, *if p then not q*, *if not p then q* and *if not p then not q*. The instructions were modeled on those used by Evans, Clibbens, and Rood (1996), and a full set of examples is shown in Table 1. The participants were informed that the experiment was concerned with logical reasoning and that all the problems related to cards that had a capital letter on the left and a single figure number on the right. A picture of a card was provided as an example. The participants were informed that for each problem, they would be given a "rule" together with a description of four cards to which the rule applied. They were then provided with an example of the lexical content and layout of a problem as follows:

RULE:	If the letter is R then the number is not 3		
	Left side	Right side	DECISION (✓ X ?)
Card 1	R	8	_____
Card 2	R	3	_____
Card 3	T	7	_____
Card 4	G	3	_____

The participants were informed that their task was to decide for each card whether it conformed to the rule, contradicted the rule, or was irrelevant to the rule. They were asked to indicate their response with either a tick (conforms), a cross (contradicts), or a question mark (irrelevant) next to each card description. The four cards presented under each rule corresponded to the TT, TF, FT, and FF truth table cases. Four sets of the materials were produced, with card order randomized differently within each set for all the problems. Presentation order was individually randomized for each participant.

Inference Task

The participants were presented with problem booklets consisting of 32 problems, together with the putative conclusion for evaluation (see Table 2). Two problems involving different letter/number pairs were designed for each inference type (MP, DA, AC, and MT) under each rule, resulting in 32 problems presented on a separate booklet page in an order that was individually randomized for each participant. The written instructions were modeled on those used by Evans et al. (1995), with the addition of a paragraph giving extra emphasis to logical necessity. The participants were told that they would be given a test of logical reasoning and that they would reason with statements concerning imaginary letter–number pairs. They were then shown examples of conditional statements of the kind to be used, including negated components. They were told that their task was “to decide whether or not the conclusion *necessarily* follows from the statements. A conclusion is necessary if it must be true, given that the statements are true.” They were then shown the following example of a modus ponens problem in the form used in the task:

If the letter is B then the number is 7
 The letter is B
 CONCLUSION The number is 7 YES NO

The participants were then instructed as follows:

If you think the conclusion necessarily follows then please circle the YES response; otherwise, circle the NO response. Remember, a conclusion is necessary, if it must be true, given that the statements are true. If and only if you judge that a conclusion necessarily follows from the statements, you should answer “yes”; otherwise, “no.”

Probability-of-Conditionals Task

The participants were informed that the purpose of this part of the experiment was to learn about the way in which people interpret conditional sentences. They would be presented with a series of scenarios followed by a statement, and they would then be asked to rate how likely it is that the statement is true. Sixteen problems were presented, based on the design used by Oberauer and Wilhelm (2003). For each problem, the participants were told that a pack (deck) contained cards that had one of two shapes on one side and one of two colors on the other. They were then given the frequencies of the four possible combinations of a shape with a color and were asked how likely it was that the following claim would be true of a card drawn at random from the pack. An example of the problem format used is as follows:

A pack contains cards which are either blue or green and have either a circle or a triangle printed on them. In total there are
 900 blue circles
 100 blue triangles
 500 green circles
 500 green triangles
 in the pack.
 How likely is the following claim to be true of a card drawn at random from this pack?
 If the card is blue then it has a circle printed on it

Table 3
Frequencies of Truth Table Cases in the Probability-of-Conditionals Tasks

Cases	HH	HL	LH	LL
pq	900	900	90	90
p-q	100	900	10	90
-pq	500	100	950	910
-p-q	500	100	950	910

Note—HH, high frequency of pq, high $P(q|p)$; HL, high frequency of pq, low $P(q|p)$; LH, low frequency of pq, high $P(q|p)$; LL, low frequency of pq, low $P(q|p)$.

The participants indicated their response on a scale of 1 (being *very unlikely*) to 5 (being *very likely*; taken from Evans et al., 2003). The frequencies of truth table cases given were designed to vary the conditional probability $P(q|p)$ and the frequency of pq orthogonally (following Oberauer & Wilhelm, 2003), as is illustrated in Table 3. On the basis of the four sets illustrated in the table, 3 more items of each set were constructed by dividing the original frequencies by 2, 5, and 10, resulting in 4 items of each type and 16 in all. Note that for simplicity and in contrast with the previous tasks, only affirmative conditional statements were used. Our previous research suggests that use of negated components makes little difference on this task (Evans et al., 2007; Over et al., 2007).

RESULTS

The main interest in this article is the observation of qualitative and quantitative individual differences that may be revealed by observing participants’ performance across the different tasks and in relation to the general intelligence measure. However, we will first present a separate brief report on each of the three main cognitive tasks used, with appropriate analyses to establish that the normally observed trends are present.

Truth Table Task

The mean percentages of truth table classifications for all 120 participants are shown in Table 4A. On the basis of previous research, we would expect two main trends. The first is an overall tendency toward classification of logical cases in conformity with the defective truth table pattern. The data reported in Table 4A conform to this expectation, with the modal classifications being true for TT cases, false for TF cases, and irrelevant for both FT and FF cases. However, it can also be observed that there is a great deal of variability of classification within cases other than TT. This too is normally observed, and the main reason that this is likely to happen is matching bias. Again, the data look normal, since for all four cases, judgments of irrelevance are highest for double-mismatch cases (shown in Table 4A as --) and lowest for double-match cases (++) , with single-match cases intermediate.

Following previous practice (e.g., Evans et al., 1996a), we computed two indices for each individual participant to measure the matching bias effect with a possible range from -8 to +8. These are measured as follows.

Antecedent matching index (AMI): frequency of irrelevant judgments on $\neg pq$ and $\neg p\neg q$ cases minus frequency of irrelevant judgements on pq and $p\neg q$ cases.

Consequent matching index (CMI): frequency of irrelevant judgments on $p\neg q$ and $\neg p\neg q$ cases minus frequency of irrelevant judgments on pq and $\neg pq$ cases.

The mean AMI and CMI scores are shown in Table 4B. As was expected, both have mean scores significantly above zero, as assessed by a binomial test ($p < .001$ in each case). For the purposes of the individual-differences analysis to follow, we also computed two further indices, with descriptive statistics again shown in Table 4B. A key prediction was that conditional probability responders would be of higher cognitive ability and show a more defective truth table pattern. We computed two separate indices of relevance to this prediction, as described in the introduction: the logic index

Table 4A
Results From the Truth Table Task Based on All Participants (N = 120): Mean Percentages of Truth Table Classifications

Rule	TT			TF			FT			FF						
	T	F	?	T	F	?	T	F	?	T	F	?				
If p then q	98	2	0	++	1	93	6	+–	6	41	53	–+	6	6	88	--
If p not-q	93	3	4	+–	2	97	1	++	10	5	85	--	26	8	66	+–
If not-p then q	91	5	4	–+	4	48	48	--	7	74	19	++	28	18	54	+–
If not-p not-q	57	8	35	--	18	68	14	+–	23	27	50	+–	36	34	30	++

Note—++ , double-matching case, pq; +– , antecedent match, p not-q; –+ , consequent match, not-pq; -- double-mismatching case, not-p not-q.

(TT as true plus TF as false) and the FAI. Each of these indices has a possible range of 0–8. As one would expect, both indices show means well above zero. We will examine their relation to cognitive ability later in the article.

It should be noted that although AMI and CMI are statistically independent, as are the FAI and logic indices, there is some partial nonindependence between the matching indices and the others. For example, if false antecedent classifications of irrelevance were very high, this would limit the matching indices, since they are measured as the difference in irrelevant classifications between matching and mismatching cases. Similarly, if logical classifications of TT and TF are very high, there is little scope for matching bias on true antecedent cases. Correlations between indices confirmed these relationships (see Table 7). The FAI showed a small but significantly negative correlation with both AMI and CMI. The logic index showed substantial negative correlations with both AMI and CMI, which suggest, perhaps, more than simply a measurement problem. We also found some significant empirical correlations between the indices that are measured independently. The two matching indices were positively correlated, as were the logic index and the FAI.

Conditional Inference Task

The frequency of endorsement of the four conditional inferences is shown in Table 5 for the whole sample, broken down by the type of conditional presented. The data are fairly typical of those observed with abstract conditionals, with near universal MP rates, high endorsements of AC, and highly variable endorsements of the *denial* inferences DA and MT. The last are expected to show a strong negative conclusion bias, since the drawing of affirmative conclusions involves resolution of a double negation (see the introduction). This pattern is confirmed (see asterisked cases). Binomial tests actually showed a highly significant negative conclusion bias for DA, MT, and AC (all at $p <$

.001, one-tailed). The last trend is sometimes observed but seems to be weaker (Schroyens et al., 2001). It is not surprising that it is significant in this experiment with the unusually large sample. For purposes of the individual-differences analyses that follow, we computed a double-negation index (DNI) by subtracting the number of DA and MT inferences endorsed when the conclusion given was affirmative from those given when it was negative. (A negative conclusion bias can be clearly interpreted as a double-negation effect only for these two inferences.)

Probability-of-Conditionals Task

One participant failed to complete this task, resulting in an n of 119 for all the analyses that follow. The ratings on all 16 problems were collated. Preliminary analyses indicated that there were no significant differences among the different versions of problems within each of the main categories: HH [high frequency of pq , high $P(q|p)$], HL [high frequency of pq , low $P(q|p)$], LH [low frequency of pq , high $P(q|p)$], and LL [low frequency of pq , low $P(q|p)$], so the data were collapsed into these four categories for all further analyses. The mean ratings in each condition are shown in Table 6. As was expected, ratings are higher both for higher pq frequency cases and higher $P(q|p)$ cases. This was confirmed by running a 2×2 repeated measures ANOVA. Both pq frequency [$F(1,118) = 78.48, MS_e = 81.74, p < .001$] and level of conditional probability $P(q|p)$ [$F(1,118) = 218.10, MS_e = 125.33, p < .001$] yielded highly significant main effects, but with no interaction.

Our main interest in this task is to demonstrate, as was observed by both Evans et al. (2003) and Oberaur and Wilhelm (2003), that these two main effects are largely produced by *different* participants. That is to say, some respond mostly on the basis of conjunctive probability [$P(pq)$] and some mostly on the basis of conditional probability [$P(q|p)$]. We computed two (orthogonal) indices on each participant as follows.

Conjunctive probability index: mean ratings for HH plus HL minus mean ratings for LH and LL.

Conditional probability index: mean ratings for HH plus LH minus mean ratings for HL and LL.

As was expected, the two indices were substantially negatively correlated ($-.788, p < .01$, two-tailed), which suggests two independent response modes. In order to demonstrate the underlying qualitative individual differences, we rank ordered all participants by their score on the conditional probability index and then split them into 10 groups. The mean for each of these groups (ordered by

Table 4B
Results From the Truth Table Task Based on All Participants (N = 120): Descriptive Statistics for the Four Indices Computed for the Truth Table Task

Index	Minimum	Maximum	<i>M</i>	<i>SD</i>
Antecedent matching	–2.00	8.00	2.30	1.79
Consequent matching	–2.00	5.00	1.82	1.51
False antecedent	0.00	8.00	4.46	2.26
Logic	3.00	8.00	6.47	1.16

Table 5
Mean Percentages of Inferences Made by All Participants (N = 120) on the Conditional Inference Task

Rule	MP	DA	AC	MT
If p then q	100	55*	74	61*
If p then not-q	99*	37	68	64*
If not-p then q	96	55*	81*	35
If not-p then not-q	95*	40	74*	39

Note—Cases in italics have affirmative premises. MP, modus ponens; DA, denial of the antecedent; AC, affirmation of the consequent; MT, modus tollens. *The argument has a negative conclusion.

Table 6
Mean Likelihood Ratings (With Standard Deviations) in the Probability-of-Conditionals Task (N = 119)

	High P(q p)		Low P(q p)	
	M	SD	M	SD
High pq frequency	4.33	0.62	3.29	0.53
Low pq frequency	3.49	1.37	2.48	0.75

conditional probability score) on both indices is shown in Figure 1. It is clear from the figure that most participants rely heavily on one cue or the other. The two main effects in the ANOVA reported above are indeed produced by different subsets of participants.

In order to test our a priori predictions, we needed to divide the participants into two groups, dependent on whether they predominantly used conditional probability or conjunctive probability as the basis for their judgments. To do this, we first normalized each index by converting it to a z score. We then assigned people to either the conditional probability group (conditional probability responders) or the conjunctive probability group (conjunctive probability responders), depending on which index was higher for each participant. A small number of participants were excluded, since they had a negative score on both indices. On this basis, we classified 46 as conjunctive probability responders and 69 as conditional probability responders, confirming earlier findings that the conditional probability mode was the more common (Evans et al., 2003; Oberauer & Wilhelm, 2003). We will report differences in performance on other tasks between these two groups below.

Individual Differences in Cognitive Ability

Our main interest in this article is in the exploration of qualitatively distinct groups of reasoners, based on the difference between conditional probability and conjunctive probability responders. However, we first will present a correlation matrix (Table 7) in which the main measures of interest are included. We will comment only briefly on these where the relevant trends emerge more clearly from the comparisons of qualitatively distinct groups to follow. The correlations are of most interest with regard to general intelligence. AH4 scores correlated positively with the conditional probability index and negatively with the conjunctive probability index, as was predicted, although the latter fell short of significance. On the truth table task, high ability was associated with significantly better logi-

cal performance and less matching bias (not significant). As also was predicted, high-ability people were significantly more likely to classify false antecedent cases as irrelevant, in line with the defective truth table. This was shown by the significant correlation ($r = .25, p < .01$) of AH4 with the FAI. We also predicted negative correlations of AH4 with matching bias. Although negative, these correlations were small and nonsignificant.

Individual differences in the conditional inference task were studied by Newstead et al. (2004), who, as was noted earlier, found a surprising result in a study of comparable abstract conditional reasoning problems and also used the AH4 measure of intelligence. Higher ability individuals gave more MP inferences but fewer DA, AC, and MT inferences. The anomalous finding concerns MT, which, although a logically valid inference, is endorsed less frequently by high-ability participants, against the general trend that high-ability people reason more normatively (Stanovich, 1999). It should be noted that the apparently conflicting findings of De Neys et al. (2005) were derived from study of realistic (causal), rather than abstract, conditionals.

The overall correlations (Table 7) were in the same direction as those reported in Newstead et al. (2004) for all inferences but were significant only for MP. Following Newstead et al., we also ran an extreme group comparison of the top ($n = 34$) and bottom ($n = 32$) quartiles on the AH4 test. We predicted the following trends, each assessed with a one-tailed *t* test. We show the actual percentages of acceptance rates in parentheses:

1. More MP in high (98%) than in low (96%) ability groups, $t = 1.16$ (n.s.).
2. Less DA in high (34%) than in low (50%) ability groups, $t = 1.91$ ($p < .05$).
3. Less AC in high (64%) than in low (78%) ability groups, $t = 1.83$ ($p < .05$).
4. Less MT in high (43%) than in low (52%) ability groups, $t = 1.11$ (n.s.).

On this more powerful test, we replicated Newstead et al.'s (2004) findings for DA and AC, but not for MP

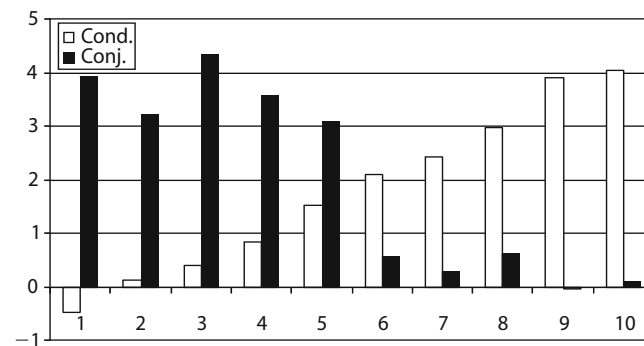


Figure 1. Mean conditional probability (Cond.) and conjunctive probability (Conj.) indices in 10 groups, ordered by Cond. score.

Table 7
Correlations Matrix Showing Relations Between AH4 Scores and Measures From the Three Cognitive Tasks

Task		AMI	CMI	FAI	Logic	MP	DA	AC	MT	DNI	COND	CONJ
Truth table	AH4	-.015	-.022	.251**	.253**	.194*	-.149	-.150	-.071	.032	.202*	-.172
	AMI		.524**	-.194*	-.599**	-.095	.198*	.159	.130	.185*	-.314**	.301**
	CMI			-.220*	-.512**	-.132	.283**	.302**	.175	.074	-.313**	.215*
	FAI				.350**	.022	-.270**	-.332**	-.245**	-.047	.230*	-.235**
Inference	Logic					.152	-.243**	-.287**	-.186*	-.100	.290**	-.269**
	MP						-.012	.186*	.110	.109	.116	-.140
	DA							.559**	.823**	.077	-.158	.138
	AC								.413**	.118	-.202*	.184*
	MT									.119	-.039	-.003
Probability of conditionals	DNI										.061	-.014
	COND											-.788**
	CONJ											

Note—AMI, antecedent matching index; CMI, consequent matching index; FAI, false antecedent index; MP, modus ponens; DA, denial of the antecedent; AC, affirmation of the consequent; MT, modus tollens; DNI, double-negation index; COND, conditional probability index; CONJ, conjunctive probability index. *Correlation significant at .05 level (two-tailed). **Correlation significant at .01 level (two-tailed).

and MT. Note, however, that the comparisons above were based on inference rates aggregated across the four types of conditional statements. For strict comparability to Newstead et al., we should consider affirmative conditionals only. These analyses led, however, to very similar findings. Both DA rates ($t = 2.63, p < .01$) and AC rates ($t = 2.81, p < .01$) were significantly lower in the high-ability group, whereas differences in MP and MT rates were small and nonsignificant. It is worth noting the nonsignificant trend for MT rates to be lower for high-ability participants in this study, consistent with Newstead et al. So we can at least confirm that higher ability participants do not show the advantage in MT reasoning that might be expected from the general relationship between cognitive capacity and normatively correct responses.

Like the matching indices, the DNI failed to correlate significantly with the AH4 score. Hence, we have no evidence in this study that higher ability participants are better able to resist the dominant cognitive biases in conditional reasoning tasks.

Comparison of Conditional and Conjunctive Probability Responders

Our predictions of qualitative individual differences were framed in terms of expected differences between conjunctive and conditional probability responders on the other cognitive tasks. Having divided most of the participants into one of these two groups (see above), we then compared their performance on the key measures from other tasks, as shown in Table 8. In view of the number of significance tests, we used two-tailed *t* tests throughout to check for the reliability of differences, even though most are in the predicted direction. As was predicted, conditional probability responders were of higher general intelligence and gave more accurate classifications of TT and TF cases (logic index) and more judgments of irrelevance on FT and FF cases (FAI), in line with the defective truth table. Also as was predicted, the conjunctive probability responders accepted more AC inferences on the conditional inference task. Hence, all our predictions about the differences between conditional and conjunctive probability responders are confirmed on the basis of our assumption that the for-

mer, but not the latter, process the conditional in a fully suppositional manner. The association of conditional probability responders with more judgments of false antecedent cases as irrelevant is particularly critical in the theoretical debate between the suppositional and the mental model theory of conditionals, as was indicated above.

In addition to testing these prior predictions, we also tested for differences in susceptibility to cognitive biases. We have already shown that these were uncorrelated with AH4 scores. Despite the presence of a strong double-negation effect on the conditional inference task, there was similarly no evidence of a difference in the strength of the double-negation effect in the analysis between conjunction probability and conditional probability responders. However, matching bias was significantly reduced in conditional probability responders on both AMI and CMI measures. Presumably, this reflects some difference in processing style between these two groups that

Table 8
Differences Between Conjunctive and Conditional Probability Responders on General Intelligence and Performance Measures Derived From Other Tasks

Measure	Conjunctive Probability (N = 46)		Conditional Probability (N = 69)		<i>t</i>	Significance
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Intelligence						
AH4	91.65	14.29	98.07	12.67	-2.45	.013
Truth table task						
AMI	2.88	1.69	1.74	1.59	3.62	.000
CMI	2.23	1.50	1.54	1.47	2.44	.016
FAI	3.85	1.88	4.95	2.43	-2.53	.011
Logic	6.16	1.00	6.76	1.17	-2.80	.005
Inference task						
MP	3.86	0.40	3.92	0.20	-1.12	n.s.
DA	2.12	1.18	1.67	1.52	1.61	n.s.
AC	3.26	1.06	2.72	1.42	2.09	.031
MT	2.06	1.18	1.94	1.35	0.37	n.s.
DNI	0.81	1.01	0.81	1.00	0.02	n.s.

Note—AMI, antecedent matching index; CMI, consequent matching index; FAI, false antecedent index; MP, modus ponens; DA, denial of the antecedent; AC, affirmation of the consequent; MT, modus tollens; DNI, double-negation index.

is not related to their difference in general intelligence. It also means that the main cause of irrelevant judgments differs between the two groups. Conditional probability responders are more likely to see false antecedent cases as irrelevant but less likely to see mismatching cases as irrelevant, an entirely new and striking result. The FAI is also significantly negatively correlated with both the AMI and the CMI (see Table 7).

DISCUSSION

In this experiment, we presented each of several tasks involving the processing of abstract indicative conditional statements to single large group of participants. Our results are unambiguous in supporting our main theoretical predictions about the differences between people who judge the probability of a conditional on the basis of conditional or conjunctive probability. On the assumption that conditional probability, but not conjunctive probability, responders are processing the conditional statement in a fully suppositional manner, we predicted that the former would be (1) of higher cognitive ability and (2) more likely to classify conditionals according to the defective truth table on the truth table task. These predictions were strongly supported. In addition, we are able to show directly that higher ability participants are more likely to give defective truth table responses. For reasons explained earlier, this finding strongly distinguishes the suppositional and the mental model theory accounts of the conditional. Defective truth table responses are attributed by Johnson-Laird and Byrne (2002) to the failure to flesh out explicitly the full mental models of the conditional. Hence, they must predict that those of higher ability (and higher working memory capacity) will be less, and not more, likely to classify false antecedent cases as irrelevant.

Owing to the large resource demands of the present experiment, it was not possible to run a replication experiment. However, we recently learned of a closely related study by Oberauer et al. (2007, Experiment 3) that was designed, like ours, to explicate the basis of the difference between people who base the probability of conditionals on conditional or conjunctive probabilities. For this reason, we will comment in some detail on their study and the relation of their findings to ours. Like us, they administered the probability-of-conditionals, truth table, and conditional inference tasks to the same participants on a slightly smaller sample ($N = 100$, rather than 120). The other differences between the two studies are the following: (1) Oberauer et al. did not administer a test of general intelligence or any other measure of cognitive capacity; (2) they applied the negations paradigm to the probability-of-conditionals task (we did not); (3) they did not apply the negations paradigm to the conditional inference task (we did); (4) their participants received a truth table task similar to the one we administered but were also given a task with no *irrelevant* response and were forced to choose between true and false. Both studies used the negations paradigm with the truth table task. There were several differences in the methods of analysis used in the two studies. In particular, Oberauer

et al. analyzed holistic truth table patterns, whereas we abstracted indices, for reasons explained below. For comparability of the two studies, however, we computed similar classifications, which are compared with the data of Oberauer et al. in the Appendix. Note that the two studies appear quite consistent. The correlation across the 64 cells between the two data sets is .86.

Our own study bears on the issue of how conjunctive and conditional probability responders differ in their general thinking about conditional statements, but also on wider issues having to do with individual differences in cognitive ability. We will focus this discussion initially on the first issue, where we can make direct comparisons with the findings of Oberauer et al. (2007). Both studies replicated earlier findings with the probability-of-conditionals tasks, so that both probabilities affected responding, with the effect of conditional probability stronger. Oberauer et al. found little effect of negations on the judgment of probability-of-conditional statements (as did Over et al., 2007). Both studies were also able to classify individual participants as predominantly conditional probability or conjunctive probability responders. We classified 58% as conditional probability responders and 38% as conjunctive probability responders, with 14% unclassified; using a slightly different criterion, they classified 70% as conditional probability responders and 18% as conjunctive probability responders, with 12% unclassified. Since we now know that these categories are linked to cognitive ability, this may well account for the different proportions observed in the two studies (i.e., Oberauer et al. may have studied a higher ability population).

The present study confirmed our predictions about the difference between conditional probability and conjunctive probability responders (Table 8), so it is interesting to check whether the same was true in the data of Oberauer et al. (2007) for the inference and truth table tasks. On the inference task, they did not report fully the inference rates for the two groups separately, commenting instead that the only significant difference was that fewer DA inferences were made in their conditional probability group (58%) than in their conjunctive probability group (74%). The corresponding rates in our study were 42% and 53%, a trend that fell short of significance. Oberauer (personal communication, December 2006) reports that although they did not find a significant difference for AC inferences, the trend was for more such inferences to be made by the conjunctive probability group (70%) than by the conditional probability group (61%), consistent with the significant trend in the same direction (82% conjunctive probability group, 68% conditional probability group) that we have reported.

Oberauer et al. (2007) concentrated most of their individual-differences discussion on the truth table task, where they took an approach different from ours—classifying individuals according to their whole truth table pattern. We avoided this method (also used by some earlier researchers) because we believed that matching bias operates as a response bias, which confuses any attempt to read individual truth tables as a determinate pattern (but see the Appendix). Considering their comparable truth table

task (with true, false, and irrelevant responses allowed), there is clear corroboration of one our most important findings: The defective truth table pattern TFII (TT as true, TF as false, FT as irrelevant, FF as irrelevant) correlated positively (.26) and significantly with what we are calling the conditional probability index. That is, people giving defective truth tables were more likely to base their probability judgments on the conditional probability of q given p . For comparability, we correlated similar holistic classifications of defective truth tables with the conditional probability index in our own study, with almost identical results ($r = .27, p < .01$).

It seems clear, then, that conditional probability responders, as was predicted by us (but in conflict with model theory), are both of higher cognitive ability and more likely to provide defective truth tables when classifying logical cases for conditional statements. The last two are also directly linked to each other, since both the logic and the FAI indices are significantly correlated with AH4 scores (Table 7). However, this leaves us with a theoretical problem: If conjunctive probability responders are not processing the conditional in a suppositional manner, what exactly are they doing? What causes lower ability participants to give conjunctive probability responses, less defective truth tables, and more AC and DA inferences?

Considering first the truth table task, if conjunctive probability responders are less defective in their choices, it follows that they must be giving more determinate classifications of FT and FF cases. As was mentioned in the introduction, two possible reasons could be that they treat the statement as *biconditional*, in which case they would tend to classify FT as false and FF as true, or as *conjunction*, in which case they would classify both cases as false. The relevant data are shown in Figures 2 and 3. We have already shown that conditional probability responders make significantly more irrelevant classifications of these cases than do conjunctive probability responders, and Figures 2 and 3 reveal clearly what conjunctive probability responders do instead. There is no difference between the groups in true responses but a substantial shift toward false responses on both FT and FF cases for the conjunctive probability group, consistent with the truth table for conjunction: p and q . The effect is relative, since the irrelevant classification remains the modal response in both cases, showing the power of the *if* heuristic. Nevertheless, it is consistent with our hypothesis of shallow processing and a strong focus on the pq or TT case in this lower ability group who assign $P(pq)$ to the probability of the conditional statement.

What, then, of the conditional inference task? Do not the differences between the two groups suggest a biconditional reading in the conjunctive probability group, since both this group and the low-ability group tend to endorse more fallacious DA and AC inferences? We suggest that *biconditional reading* is, in fact, too grand a term for what this group is actually doing. In this case, we believe that shallow processing and a focus on the pq case results in what we call *simple equivalence* reasoning. Essentially, these participants expect p and q (or more accurately, true antecedent and true consequent) to go together. If you have one, you have the other (MP, AC); if you do not have one,

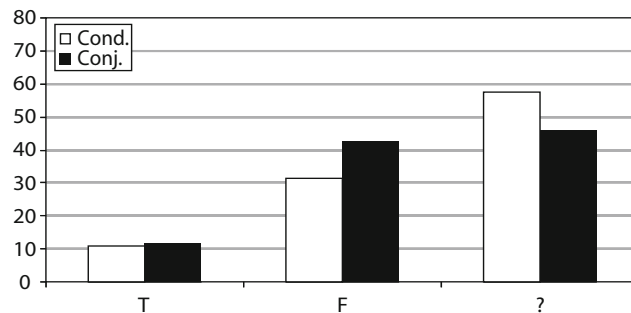


Figure 2. Truth table choices for the FT case broken down by conditional probability (Cond.) and conjunctive probability (Conj.) groups.

you do not have the other (DA, MT). Note that this reasoning strategy is symmetrical and nonsuppositional, consistent with the less suppositional processing of conditional statements by these participants on the other two tasks. If this simple equivalence strategy were the whole story, we would expect conjunctive probability participants to endorse all inferences. In practice, some inferences will actually be inhibited by the double-negation effect, as was discussed below. However, a trend for conjunctive probability participants to reason in this way will result in more DA and AC inferences in this group, relative to the conditional probability group, as we have observed, although the DA comparison fell short of significance (see Table 8). We would not, however, expect more MP inferences in the conjunctive probability group, because that is a valid inference that conditional probability reasoners should endorse, due to their correct suppositional understanding of the conditional. We will comment on the case of the MT inference shortly.

An advantage of this account is that it can explain the universally observed double-negation effect on the DA and MT inferences, which this study has shown to be independent of cognitive ability. The presence of this bias is the main reason that conjunctive probability responders do not endorse all inferences. Previously we have suggested that the double-negation effect arises in the course of suppositional reasoning (Evans & Over, 2004). However, there is little reason to think that lower ability participants and the

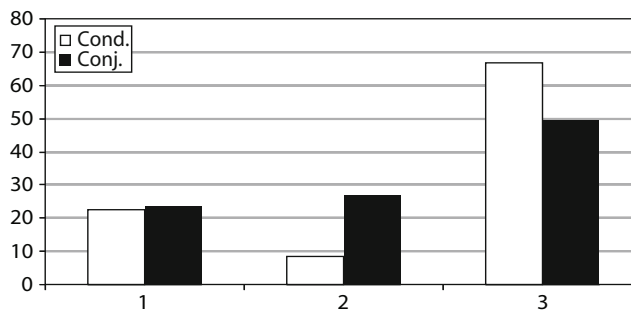


Figure 3. Truth table choices for the FF case broken down by conditional probability (Cond.) and conjunctive probability (Conj.) groups.

conjunctive probability group are reasoning suppositionally on the other tasks we presented. Hence, we now propose that the double-negation effect will also arise in the use of the simple equivalence-reasoning strategy described above. Consider the case of a DA inference, such as

If the letter is A then the number is 4
The letter is not A
Therefore, the number is not 4

These participants will reason that since you do not have the A, you cannot have the 4. Now add a negation in the consequent as follows:

If the letter is A then the number is not 4
The letter is not A
Therefore, the number is 4

Here, they reason that since you do not have an A, you cannot have a not-4. That is a double negation and one from which the conclusion will not follow without some further mental processing. Hence, we can explain the prevalence of the double-negation effect in participants who appear not to process the conditional in a suppositional manner.

What of the MT inference? If we are right, high-ability participants can complete this only by a difficult process of suppositional reasoning: a *reductio ad absurdum* of the kind described by Evans and Over (2004) and by mental logic theorists such as Braine and O'Brien (1998). Low-ability participants, by contrast, can make the inference by this much simpler and nonsuppositional strategy. Hence, they will find the inference no harder and, quite possibly, easier to make. This enables us to explain the paradoxical finding of Newstead et al. (2004) that MT inference rates were *negatively* correlated with ability, as well as the present findings that high-ability participants were no better at MT reasoning.

Throughout this article, we have talked about individual differences in the manner in which people *process* conditional statements that are presented with arbitrary materials and without context. We think it unlikely that anyone would have a mental representation of a conditional statement as a conjunction, certainly when conditionals are used in realistic, everyday contexts, because hypothetical thinking is fundamental in mental life and the conditional statement the most natural use of language to encourage its use. Hence, we believe that shallow processing of conditional statements emerges as a function of both lower cognitive ability and abstract content. In support of this argument, Over et al. (2007) have reported more widespread use of the conditional probability, $P(q|p)$, when participants are asked to judge the believability of everyday causal conditionals and no significant trend with such materials toward use of the conjunctive probability $P(pq)$.

In conclusion, this study has provided a number of novel findings that advance our understanding of the psychology of conditionals. We have confirmed that there are, indeed, two different groups of people who respond on the basis of conditional and conjunctive probability

when asked to the judge the likelihood of an abstract conditional statements, as has been suggested by earlier studies (Evans et al., 2003; Oberauer & Wilhelm, 2003). We have demonstrated that membership in these groups, in turn, corresponds with a number of significant differences on measurable performance on the conditional inference and truth table tasks, as well as in general cognitive ability, most of which were predicted a priori on the basis of our suppositional theory of conditionals. In particular, we show a three-way relationship between high cognitive ability, conditional probability responding, and defective truth tables that sharply distinguishes our account from that based on mental model theory. These findings present new challenges for psychologists interested in theoretical accounts of conditionals and demonstrate beyond doubt that such theories must be able to explain the existence of qualitative, as well as quantitative, individual differences.

AUTHOR NOTE

This research was supported by Project Grant RES-000-23-0722 from the Economic and Social Research Council of the United Kingdom. Correspondence concerning this article should be sent to J. St. B. T. Evans, Centre for Thinking and Language, University of Plymouth, Plymouth PL4 8AA, England (e-mail: jevans@plymouth.ac.uk).

REFERENCES

- BENNETT, J. (2003). *A philosophical guide to conditionals*. Oxford: Oxford University Press.
- BRAINE, M. D. S., & O'BRIEN, D. P. (1991). A theory of *If*: A lexical entry, reasoning program, and pragmatic principles. *Psychological Review*, **98**, 182-203.
- BRAINE, M. D. S., & O'BRIEN, D. P. (1998). The theory of mental-propositional logic: Description and illustration. In M. D. S. Braine & D. P. O'Brien (Eds.), *Mental logic* (pp. 79-89). Mahwah, NJ: Erlbaum.
- COLOM, R., REBOLLO, I., PALACIOS, A., JUAN-ESPINOSA, M., & KYLLONEN, P. C. (2004). Working memory is (almost) perfectly predicted by *g*. *Intelligence*, **32**, 277-296.
- EDGINGTON, D. (1995). On conditionals. *Mind*, **104**, 235-329.
- EDGINGTON, D. (2003). What if? Questions about conditionals. *Mind & Language*, **18**, 380-401.
- EVANS, J. ST. B. T. (1998). Matching bias in conditional reasoning: Do we understand it after 25 years? *Thinking & Reasoning*, **4**, 45-82.
- EVANS, J. ST. B. T. (2007). *Hypothetical thinking: Dual processes in reasoning and judgement*. Hove, U.K.: Psychology Press.
- 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., CLIBBENS, J., & ROOD, B. (1996). The role of implicit and explicit negation in conditional reasoning bias. *Journal of Memory & Language*, **35**, 392-409.
- EVANS, J. ST. B. T., ELLIS, C. E., & NEWSTEAD, S. E. (1996). On the mental representation of conditional sentences. *Quarterly Journal of Experimental Psychology*, **49A**, 1086-1114.
- EVANS, J. ST. B. T., HANDLEY, S. J., HADJICHRISTIDIS, C., THOMPSON, V., OVER, D. E., & BENNETT, S. (2007). On the basis of belief in causal and diagnostic conditionals. *Quarterly Journal of Experimental Psychology*, **60**, 635-643.
- EVANS, J. ST. B. T., HANDLEY, S. J., & OVER, D. E. (2003). Conditionals and conditional probability. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 321-355.
- EVANS, J. ST. B. T., NEWSTEAD, S. E., & BYRNE, R. M. J. (1993). *Human reasoning: The psychology of deduction*. Hove, U.K.: Erlbaum.
- EVANS, J. ST. B. T., & OVER, D. E. (2004). *If*. Oxford: Oxford University Press.
- EVANS, J. ST. B. T., OVER, D. E., & HANDLEY, S. J. (2005). Supposition,

extensionality and conditionals: A critique of Johnson-Laird & Byrne (2002). *Psychological Review*, **112**, 1040-1052.

GIROTTO, V., & JOHNSON-LAIRD, P. N. (2004). The probability of conditionals. *Psychologia*, **47**, 207-225.

HADJICHRISTIDIS, C., STEVENSON, R. J., OVER, D. E., SLOMAN, S. A., EVANS, J. ST. B. T., & FEENEY, A. (2001). On the evaluation of *If p then q* conditionals. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the 23rd Annual Meeting of the Cognitive Science Society* (pp. 409-414). London: Erlbaum.

HANDLEY, S., EVANS, J. ST. B. T., & THOMPSON, V. A. (2006). The negated conditional: A litmus test for the suppositional conditional? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **32**, 559-569.

JOHNSON-LAIRD, P. N., & BYRNE, R. M. J. (1991). *Deduction*. Hove, U.K.: Erlbaum.

JOHNSON-LAIRD, P. N., & BYRNE, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics and inference. *Psychological Review*, **109**, 646-678.

KLACZYNSKI, P. A., & DANIEL, D. B. (2005). Individual differences in conditional reasoning: A dual-process account. *Thinking & Reasoning*, **11**, 305-325.

KYLLONEN, P., & CHRISTAL, R. E. (1990). Reasoning ability is (little more than) working memory capacity?! *Intelligence*, **14**, 389-433.

MARCUS, S. L., & RIPS, L. J. (1979). Conditional reasoning. *Journal of Verbal Learning & Verbal Behavior*, **18**, 199-233.

NEWSTEAD, S. E., HANDLEY, S. J., HARLEY, C., WRIGHT, H., & FARELY, D. (2004). Individual differences in deductive reasoning. *Quarterly Journal of Experimental Psychology*, **57A**, 33-60.

OAKSFORD, M., & CHATER, N. (2001). The probabilistic approach to human reasoning. *Trends in Cognitive Sciences*, **5**, 349-357.

OBERAUER, K., GEIGER, S. M., FISCHER, K., & WEIDENFELD, A. (2007). Two meanings of "If": Individual differences in the interpretation of conditionals. *Quarterly Journal of Experimental Psychology*, **60**, 790-819.

OBERAUER, K., & WILHELM, O. (2003). The meaning(s) of conditionals: Conditional probabilities, mental models and personal utilities. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 680-693.

OVER, D. E., HADJICHRISTIDIS, C., EVANS, J. ST. B. T., HANDLEY, S. J., & SLOMAN, S. A. (2007). The probability of causal conditionals. *Cognitive Psychology*, **54**, 62-97.

RAMSEY, F. P. (1931). *The foundations of mathematics and other logical essays*. London: Routledge & Kegan Paul.

RIPS, L. J., & MARCUS, S. L. (1977). Suppositions and the analysis of conditional sentences. In M. A. Just & P. A. Carpenter (Eds.), *Cognitive processes in comprehension* (pp. 185-219). New York: Wiley.

SCHROYENS, W., & SCHAEKEN, W. (2004). Guilt by association: On *if* propositions and the proper treatment of mental models theory. *Current Psychology Letters*, **12**. Available at cpl.revues.org/document411.html.

SCHROYENS, W., SCHAEKEN, W., & D'YDEWALLE, G. (2001). The processing of negations in conditional reasoning: A meta-analytic study in mental models and/or mental logic theory. *Thinking & Reasoning*, **7**, 121-172.

SPERBER, D., CARA, F., & GIROTTO, V. (1995). Relevance theory explains the selection task. *Cognition*, **57**, 31-95.

STANOVICH, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. Mahwah, NJ: Erlbaum.

STANOVICH, K. E., & WEST, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral & Brain Sciences*, **23**, 645-726.

WASON, P. C. (1966). Reasoning. In B. M. Foss (Ed.), *New horizons in psychology I* (pp. 106-137). Harmondsworth, U.K.: Penguin.

APPENDIX
Comparison of Holistic Truth Table Patterns Observed in the Study of Oberauer, Geiger, Fischer, and Weidenfeld (2007) and Percentage Frequency Distributions in the Present Study

Pattern	Oberauer et al. Study				Present Study			
	AA	AN	NA	NN	AA	AN	NA	NN
TFFT	1	2	2	1	2	5	2	4
TFFT	4	0	3	7	4	1	8	5
TFFF	2	3	3	4	4	2	6	3
TFII	36	45	35	34	49	58	9	22
TFFI	36	4	22	4	30	5	13	1
TFIT	0	21	0	13	0	17	0	7
TIFI	9	0	13	0	3	0	22	0
TIFF	3	1	4	1	0	0	7	1
TIFT	1	0	7	0	0	0	14	0
TFTF	1	1	1	0	0	0	0	0
TTFT	0	0	0	0	0	0	2	0
FTTF	0	0	0	1	0	0	0	1
FTIF	0	5	1	2	0	0	0	1
FTII	0	2	0	6	0	1	0	1
IFIT	0	5	0	2	0	1	0	2
IFFT	0	2	0	9	0	0	0	12
Others	9	12	12	17	8	10	17	40

Note—AA, if p then q; AN, if p then not-q; NA, if not-p then q; NN, if not-p then not-q. T, true; F, false; I, irrelevant. The four letters in a pattern refer to judgments of the four truth table cases in the order TT, TF, FT, and FF. Percentages for the present study are derived from responses on two problems of each conditional type (AA, AN, NA, NN), for a total of 120 participants.

(Manuscript received September 20, 2006;
 revision accepted for publication December 16, 2006.)