Journal of Experimental Psychology:
Learning, Memory, and Cognition
2011, Vol. 37, No. 3, 635–648

0278-7393/11/$12.00 DOI: 10.1037/a0022329

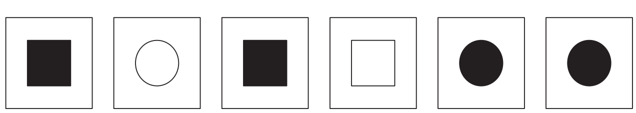
How People Interpret Conditionals: Shifts Toward the Conditional Event

Andrew J. B. Fugard, Niki Pfeifer, Bastian Mayerhofer, and Gernot D. Kleiter
University of Salzburg

We investigated how people interpret conditionals and how stable their interpretation is over a long series of trials. Participants were shown the colored patterns on each side of a 6-sided die and were asked how sure they were that a conditional holds of the side landing upward when the die is randomly thrown. Participants were presented with 71 trials consisting of all combinations of binary dimensions of shape (e.g., circles and squares) and color (e.g., blue and red) painted onto the sides of each die. In 2 experiments ($N_1 = 66$, $N_2 = 65$), the conditional event was the dominant interpretation, followed by conjunction, and material conditional responses were negligible. In both experiments, the percentage of participants giving a conditional event response increased from around 40% at the beginning of the task to nearly 80% at the end, with most participants shifting from a conjunction interpretation. The shift was moderated by the order of shape and color in each conditional's antecedent and consequent: Participants were more likely to shift if the antecedent referred to a color. In Experiment 2 we collected response times: Conditional event interpretations took longer to process than conjunction interpretations (mean difference = 500 ms). We discuss implications of our results for mental models theory and probabilistic theories of reasoning.

Keywords: conditionals, understanding conditional probability, stability of interpretation, change points, probabilistic truth table task

The conditional in all its forms has received much attention across the disciplines as it is ubiquitous in inference, for instance in conversation, hypothetical thinking when solving problems, and in inferences about inference in mathematical logic. Not only are there many theoretical approaches to explaining how people understand and reason about conditionals, there are also interindividual differences in how people understand *if*. To illustrate the different interpretations, consider a fair die with the following patterns on the sides:



The die is thrown randomly and lands with one side facing upward. What is the truth value of the following sentence (if indeed it has a truth value):

> If the side shows a square, then the side shows black?

Most people agree that if the side shows a black square, then the conditional is *true*, and if the side shows a white square, then the conditional is *false*. Opinions diverge when the antecedent (i.e., "the side shows a square") is false. One response is that the conditional is *false* (the *conjunction* interpretation), another is that it is *true* (the *material conditional* interpretation), and yet another is that the conditional is neither true nor false (what we will call the *conditional event* interpretation, more on which later). See Table 1 for a summary of these interpretations.

Suppose we don't know which side of the die landed facing upward. What can be said about the conditional? Now it is not possible to have a certain belief about whether the sentence is true or false, but there is information that can allow one to infer a degree of belief in the conditional. How sure can you be that the above conditional holds? If you take a probabilistic approach to the problem, then still there are three answers. For the conjunction interpretation, you consider how many cases there are when the conditional could be true (i.e., showing a black square) and divide by the total number of possibilities, (i.e., 6), giving for this die and conditional the probability 2/6 or 1/3. For the material conditional interpretation, what matters are the cases where the conditional is not false, i.e., not showing a white square; again dividing by the total number of possibilities gives 5/6. Finally for the conditional event interpretation, the cases where the antecedent is false are seen as irrelevant and not counted (note the connection with the case where it is known with certainty what side is facing upward, discussed above). To infer the probability, again the number of cases where the conditional is true is counted, but now they are

Andrew J. B. Fugard, Niki Pfeifer, Bastian Mayerhofer, and Gernot D. Kleiter, Department of Psychology, University of Salzburg, Salzburg, Austria.

Niki Pfeifer is now at the Munich Center for Mathematical Philosophy, LMU Munich, Germany. Bastian Mayerhofer is now at the Courant Forschungszentrum "Textstrukturen," Georg-August-Universität Göttingen, Germany.

Parts of this work were presented at the London Reasoning Workshop (July, 2009) and the 8th Workshop on Uncertainty Processing (September, 2009) in Liblice, Czech Republic. This research was supported by Austrian Science Fund Projects I141, within the European Science Foundation LogICCC program, and P20209. We thank Sabine Eichbauer for producing the response sheet for Experiment 1 and scanning with automatic survey software; Hans Lechner for producing the button box for Experiment 2; and Pierre Barrouillet, Shira Elqayam, Angelo Gilio, Leonhard Kratzer, David Over, and Alessandra Tasso for helpful comments.

Correspondence concerning this article should be addressed to Andrew J. B. Fugard, who is now at the ICT&S Center, University of Salzburg, Sigmund-Haffner-Gasse 18, 5020 Salzburg, Austria. E-mail: andy.fugard@sbg.ac.at

Table 1
Truth Tables of the Material Conditional ($A \supset B$), Conjunction ($A \wedge B$), and Semantic Values of the Conditional Event ($B|A$)

A	B	$A \supset B$	$A \wedge B$	$B\|A$
true	true	true	true	true
true	false	false	false	false
false	true	true	false	void
false	false	true	false	void

divided by cases where the antecedent is true (i.e., those sides showing a square), giving 2/3.

Until the late 1990s, the majority of psychological theories of conditional reasoning have used classical logic as the framework for competence and performance models. The *mental rules* and *mental logic* theories stem from natural deduction systems for classical logic (Braine & O'Brien, 1998; Rips, 1994). According to these theories, when people reason about conditionals they construct mental representations of proofs. Errors in reasoning may be due, for example, to limited working memory capacity for representing and constructing the proofs or to a failure to recognize the applicability of inference rules in a particular situation. Although the classical material conditional is at the core of these theories, not all classically valid conclusions are predicted to be drawn. The mental logic theory of *if* by Braine and O'Brien (1991), for instance, has constraints on the rules that apply, resulting in a conditional with a false antecedent being seen as irrelevant.

The theory of *mental models* stems from a fragment of model theory of classical logic (Johnson-Laird, 1983; Johnson-Laird & Byrne, 2002). According to the theory, reasoners represent the possibilities in which a sentence is true. If a possibility is absent from the representation, then the sentence is considered false for that possibility. There are three interpretations of conditionals that are as independent as possible from background knowledge predicted by the theory. Two interpretations result from a partial representation. For the conditional "If the side shows a square, then the side shows black," would be represented as an explicit single model (possibility) *square* ∧ *black* and an implicit model indicating that there are other unrepresented possibilities. The interpretation given for this interpretation is the conjunction. It is also possible according to the theory that "naive individuals and theorists" (Johnson-Laird & Byrne, 2002, p. 657) see the conditional as irrelevant when its antecedent is false because of the absence of a mental model representing the false antecedent cases. The fleshed-out set of models, which in the theory gives the core meaning of the indicative conditional, explicitly represent in addition the cases ¬*square* ∧ *black* and ¬*square* ∧ ¬*black* (¬ represents negation). The case square ∧ black is not represented, so the conditional is false for that possibility. This leads to the material conditional interpretation of *if*. The theory predicts that when a task is not demanding, the representation can be fleshed out to the fully explicit one with three models (Johnson-Laird, Byrne, & Girotto, 2009). It also predicts that those with better working memory capacity would be more likely to give this interpretation of the conditional.

An alternative view gaining in popularity in psychology is that the semantics of an indicative "if A, then B" is given by the conditional probability, $P(B|A)$ (Evans, Handley, & Over, 2003; Oaksford & Chater, 2007, 2009; Pfeifer & Kleiter, 2005a, 2005b, 2009) This interpretation can be traced back several decades in philosophy (see, e.g., Adams, 1975; Bennett, 2003; Edgington, 1995). One basic advantage of a probabilistic semantics is that conditionals are no longer only true or false but can be assigned a degree of belief. This intuitively matches how we use conditionals, especially when making predictions about events that have not yet occurred.

One problem with classical logic is that it is monotonic: Conclusions cannot be withdrawn if new information is learned. This property is psychologically implausible as people often revise previously drawn conclusions (Byrne, 1989; Stenning & van Lambalgen, 2005). A problem of the material conditional interpretation is that it is truth-functional, a consequence of which are the so-called paradoxes of the material conditional. The material conditional is true if its antecedent is false or if its consequent is true. This implies that "If Bill Gates is bankrupt, then Bill Gates is a billionaire" follows logically from "Bill Gates is a billionaire." This is paradoxical. As discussed by Evans, Over, and Handley (2005), both problems disappear in a probability semantics (see Pfeifer & Kleiter, 2006, for a formal analysis). The semantics characterizes how people actually reason about the paradoxes (Pfeifer & Kleiter, 2011).

Probabilistic semantics for *if* stem from Ramsey (1929/1990), who argued that when people infer their degree of belief in "if *A*, then *B*," they assume *A*, and "fix their degrees of belief" in *B* (p. 155). If the antecedent *A* turns out to be false, "these degrees of belief are rendered *void*" (p. 155). A psychological implementation of this "Ramsey test" has been proposed (Evans, Handley, & Over, 2003, p. 325) in which people compare $P(A \wedge B)$ with $P(A \wedge \neg B)$. For the example introduced above, this would involve a comparison of the probability that a black square landed facing upward with the probability that a white square landed facing upward. According to this account, conjunction responses are due to a partial execution of the test because of limited working memory or insufficient motivation. And indeed, Evans, Handley, Neilens, and Over (2007) found evidence that those with a higher score on a composite measure of cognitive ability, interpreted as reflecting working memory capacity, were more likely to give a conditional event than a conjunction interpretation.

The traditional approach to probability defines an unconditional probability function from sentences expressed in two-valued classical propositional logic, that is, sentences that can be either true or false, to a value in the interval 0 to 1. This allows the probability of sentences such as "The side shows a square" or "The side shows a square *and* the side shows black" to be inferred. Conditional probabilities, $P(B|A)$, are defined in terms of unconditional probabilities using the ratio formula:

$$P(B|A) \overset{\text{def}}{=} \frac{P(A \wedge B)}{p(A)}, \text{ if } p(A) > 0$$

This notation is misleading as $B|A$ does not exist independently of the probability function, whereas in this traditional approach it is possible to reason about *A*, *B*, and, for example, conjunctions of *A* and *B* for the cases when the events' statuses are known with certainty, independently of their probability. Another alternative approach introduced by de Finetti (1937/1980) is to define a logic where the conditional event, $B|A$, is a fully respectable logical connective and not only defined in terms of unconditional probabilities. Probability functions can be applied directly to these conditional events. The semantic values of $B|A$ are the same as those for conjunction and material conditional when A is true and *void* when A is false (see Table 1). The semantic values for the conditional event correspond to the so-called "defective truth table," better named the de Finetti table (Manktelow, Over, & Elqayam, 2011), a pattern of responses often shown by participants in reasoning experiments (Johnson-Laird & Tagart, 1969; Wason, 1966). Participants who respond with conditional probabilities are also those who show the de Finetti table (Evans et al., 2007). Taken together, this is strong evidence that the approach introduced by de Finetti (1937/1980) has psychological validity. The explanation for the overlap in interpretations on truth table tasks and probabilistic truth table tasks is that they both tap into the core meaning of the conditional as a conditional event.

Probabilistic approaches are often set into opposition with logical approaches. Oaksford and Chater (2009) write, "Logic provides a calculus for certain reasoning—for finding conclusions which follow, of necessity, from the premises given" (p. 72). They continue, "There has, in short, been a 'probabilistic turn' across a broad range of domains—a move away from the attempt to apply logical methods to uncertain reasoning, and toward dealing with uncertainty by the application of probability theory" (p. 74). Although there are some notable exceptions (e.g., Da Silva Neves, Bonnefon, & Raufaste, 2002; Stenning & van Lambalgen, 2001, 2004, 2008; Over, 2009), it is rarely recognized in psychology that there is much more to logic than two-valued classical logic. In fact, there are many probabilistic logics. We propose using *coherence-based probability logic* (Coletti & Scozzafava, 2002) to model uncertain reasoning. Coherence has many advantages for psychological modeling compared to alternative approaches (Pfeifer & Kleiter, 2005b, 2009). The coherence approach defines a *deductive* consequence relation between premise and conclusion probabilities, prescribing how uncertainties are transmitted from the premises to the conclusion. Conditional events are *primitive*, following de Finetti (1937/1980), and not defined in terms of unconditional probabilities. Probabilities are conceived as *degrees of belief* rather than objective quantities. Although objective frequencies may be one source for generating degree of belief, we interpret probabilities as coherent descriptions of partial knowledge states. There are also extensive mathematical results building bridges between the coherence approach and other logics that were developed independently, such as basic nonmonotonic reasoning System P (Gilio, 2002). There is also empirical evidence supporting the psychological plausibility of this semantics (Pfeifer & Kleiter, 2005b, 2010).

To apply logic to the study of cognition, the processes of reasoning *to* interpretations and *from* interpretations must be separated (Stenning & van Lambalgen, 2008). Reasoning to an interpretation requires (a) a formal language to be chosen, (b) a semantics to be assigned, and (c) a characterization of when an argument is valid (Stenning & van Lambalgen, 2008, p. 25). Once these choices have been made, then reasoning from the fixed interpretation, that is, *derivation*, may proceed. From this viewpoint, errors in reasoning are due either to mismatches in interpretation (e.g., between experimenter and participant) or a failure of derivational processes. At the computational level of analysis (Marr, 1982), the main interpretational problem in probabilistic tasks is to infer whether the probability of "if A, then B" is that of (a) a conditional event ($B|A$), (b) a conjunction ($A \wedge$ B), or (c) a material conditional ($A \supset B$). A conjunction response may also result from a mapping of the natural language if-then to a conditional event, but with the task of inferring when the conditional receives the truth value *true* (Edgington, 2003). To see this, consider the die and conditional introduced earlier, "If the side shows a square, then the side shows black." This conditional is true exactly when a black square faces upward, so asking about the probability that this conditional is *true* can be interpreted as asking the probability that both the antecedent and consequent are true. This is equivalent to P ($square \wedge black$), i.e., the conjunction rather than the conditional event probability.

The standard task for investigating how people interpret indicative uncertain conditionals is the *probabilistic truth table task* (Evans, Handley, & Over, 2003; Oberauer & Wilhelm, 2003). In this task, the joint frequency distribution is provided (i.e., frequencies of conjunctions), for example,

1 yellow circle

4 yellow diamonds

16 red circles

16 red diamonds

Participants are asked to assess how sure they are that a conditional is true of a card randomly drawn from the pack. For instance, "If the card is yellow, then it has a circle printed on it." The tasks allow the experimenter to infer how the participants interpret the conditional. Overall, studies using probabilistic truth table tasks have found that just over half of participants responded with the conditional event interpretation and the remainder responded with a conjunction interpretation (Evans, Handley, & Over, 2003; Oberauer & Wilhelm, 2003). We hypothesize that one reason for the nonnegligible minority of conjunction responses may be due to the fact that conjunctive frequencies are already presented in the task. Little support has been observed for the material conditional interpretation, though Schroyens, Schaeken, and Dieussaert (2008) managed to increase material conditional interpretations using a priming paradigm.

To date, no mental rule or logic theories characterize how people reason about uncertain conditionals (that is not to say that it is impossible to do so). Mental models theory has been extended to deal with probabilities (Girotto & Johnson-Laird, 2004; Johnson-Laird, Byrne, & Girotto, 2009; Johnson-Laird, Legrenzi, Girotto, Legrenzi, & Caverni, 1999). If events are equiprobable, for example, the sides of a fair die landing facing upward, then all possibilities are represented. For the previously introduced die example, this representation would consist of the six cases: $square \wedge black$, $\neg square \wedge \neg black$, $square \wedge black$, $square \wedge \neg black$, $\neg square \wedge black$, $\neg square \wedge black$. The conjunction interpretation is computed as the ratio of cases where $square \wedge black$ is true, divided by the total number of possible events, in this case 6. The material conditional interpretation comes from the "complete strategy" of counting those cases where $square \wedge \neg black$ is not the case. The theory suggests that those giving a conditional event interpretation "may transform a problem in a subtle way" (Girotto & Johnson-Laird, 2004, p. 208) because the antecedent is a subordinate clause. Girotto & Johnson-Laird (2004) argue that "a question of the form: 'What is the probability that if A, then C?' is readily re-interpreted as: 'If A, then what is the probability of C?'"

(p. 208). The representation constructed consists only of those cases where the antecedent is true. The number of *square* ∧ *black* cases are counted and divided by the total number in this subset representation. It has been pointed out previously that this transformation may be viewed as a version of the Ramsey test (Politzer, 2007).

The present study extends previous research on uncertain conditionals by (a) presenting the task material graphically, without using numerals, so that participants can count any frequency they require for their interpretation of the conditional; (b) not priming a representation in terms of joint frequencies; (c) presenting a long series of systematically enumerated items; (d) asking for responses of the form "*x* out of *y*" to reduce the burden of arithmetic and ease discrimination between interpretations; (e) studying response times; (f) investigating the time-course of interpretation within-participants, for instance whether there are any shifts of interpretation; and (g) studying facilitation effects of entity-first versus feature-first conditionals.

Task Development

We developed a task concerning six-sided dice, using colored shapes on each side of a given die rather than the usual dots. These patterns were varied systematically on two independent dimensions: shape (e.g., square or circle) and the shape's color (e.g., red or blue). There are 84 possible assignments of two shapes and two colors to the six sides of the dice.

The dice task and instructions were implemented in Python (Version 2.6.1; Python Software Foundation, 2008) using the Pygame (Version 1.8.1; Pygame Community, 2008) graphical library. Participants were told that the aim of the experiment was to investigate how people understand if-then sentences. It was emphasized that the dice varied between trials and that they were to reason about each independently. Three examples were also given of how the sides of the die would be represented on screen. A simple animation was shown to convey the idea of a die being placed in a cup, randomly shaken, and then the cup placed on the table so that one cannot see what side of the die shows up. Four example trials were then presented to check that the participant understood the response format. These asked how sure the participant can be that atomic sentences hold, for example, "The side shows a circle" (*Die Seite zeigt einen Kreis*).

Previous studies eliciting the production of probabilities have asked participants to respond on a 5-point scale from *very unlikely* to *very likely* (Evans et al., 2007; Evans, Handley, & Over, 2003) or probabilities on a scale from 0 to 100 (Oberauer, Geiger, Fischer, & Weidenfeld, 2007; Oberauer & Wilhelm, 2003). If participants are explicitly computing probabilities, then this requires arithmetical operations of division and rescaling, which many people find difficult (Lipkus, Samsa, & Rimer, 2001). Because we were interested in studying interpretation rather than mental arithmetic, we asked participants to respond with "*x* out of *y*" (*x aus y*). In the instructions at the beginning of the task, we presented a visual scale to explain the meaning of "out of" (see Figure 1a) and showed by example that the numerator should not exceed the denominator (though we did not use these terms).

Each test trial began with a fixation cross displayed for 1 s. Participants were shown the patterns on the sides of the die and a conditional, e.g., "If the side shows a square, then the side shows red" (*Wenn die Seite ein Viereck zeigt, dann zeigt die Seite rot*). Girotto and Johnson-Laird (2004) suggest that a question like "How sure can you be that if the side shows a square, then the side shows red?" is rephrased as, "If the side shows a square, then how sure can you be that the side shows red?" because the antecedent is a subordinate clause. This, they argue, leads to an incorrect conditional probability inference. To prevent this rephrasing, we asked the question, "How sure can you be that the following statement holds?" and the conditional appeared underneath in a box. The word "holds" (*stimmt*) was used rather than "true" (*wahr*) as used by Evans, Handley, and Over (2003) and Oberauer and Wilhelm (2003). This is to deal with the ambiguity of the word *true* raised by Edgington (2003) discussed earlier, namely that the task can be interpreted as asking the probability that the conditional receives the truth value *true*, which it does exactly when both antecedent and consequent are true: the conjunction.

Cognitive Task Analysis

All reasoning tasks involve premises and a conclusion, but what exactly are they in the dice task? The instructions communicate that the die is six sided, fair, and thrown randomly and that the probability of a side landing facing upward is 1/6. Probabilities are obtained by counting the relevant joint or marginal frequencies (i.e., the frequencies of the conjuncts). The conclusion is a natural language conditional that must be interpreted before its probability can be inferred. Table 2 shows how the chosen interpretation determines which premises may be relevant and how the presented

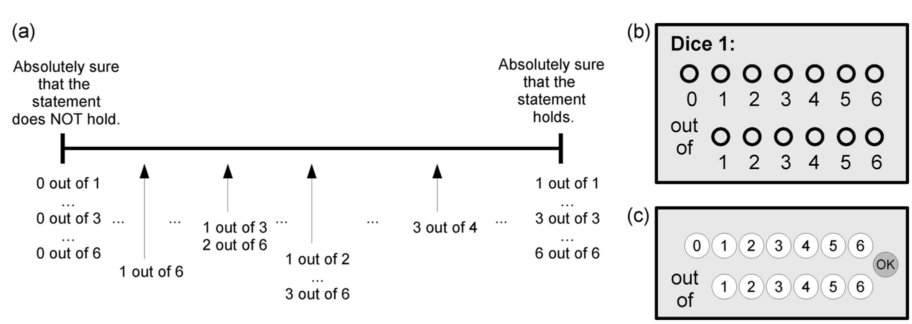


Figure 1. (a) Diagram used to convey the meaning of "out of." (b) Example of item response format on answer sheet for Experiment 1. (c) Diagram of button box for Experiment 2.

Table 2
Examples of Premises Obtainable From the Dice Presentations and How They May Be Used to Infer the Probability of the If-Then Conclusion According to the Three Interpretations

		$\lvert\neg A\rvert = f_2$
$\lvert A \wedge B\rvert = f_1$	$\lvert A \wedge B\rvert = f_1$	$\lvert A \wedge B\rvert = f_1$
$\lvert A\rvert = f_2$	$\lvert Sides\rvert = 6$	$\lvert Sides\rvert = 6$
$P(B\mid A) = \frac{f_1}{f_2}$	$P(A \wedge B) = \frac{f_1}{6}$	$P(A \supset B) = \frac{f_1 + f_2}{6}$

Note. The frequency of X is denoted by $\lvert X\rvert$.

information may be used to compute the coherent probability inferences for the three predicted interpretations.

The conditional, "if A, then B," must be parsed and committed to working memory. For the conditional event and material conditional interpretations (though not conjunction), the antecedent-consequent order matters, thus this must be respected in the memory representation of the conditional. The two propositions of the conditional consist of a subject (*the side*), a predicate (*shows*) and an object (e.g., *square*, *red*). While the subject and predicate terms were held constant in our experiments, the object terms were varied, for instance "If the side shows a square, then the side shows red" (entity-feature order) versus "If the side shows red, then the side shows a square" (feature-entity order). Mental models theory would predict that the order in which s*quare* and *red* appeared in the models would depend on the order in the conditional; however, the resulting inferences from the two orders would be indistinguishable as the *types* of the entries are abstracted away in current versions of the theory. Previous experiments on verifying verbal analogies (Kleiter, 1986) found that responses were faster for the entity-feature order than the feature-entity order, suggesting that the types matter. For instance "eye is to see as ear is to hear" was verified faster than was "see is to eye as hear is to ear." For a feature, F (e.g., sees), and concrete entity, e (e.g., eye), Kleiter (1986) argued it is easier first to form a representation of e and second bind it to $F(e)$, than first to form a representation of F, and second bind it to $F(e)$.

Intuitively, when referring to an object, its shape takes precedence over its color. This is especially true of the objects used in the present study—squares, triangles, and circles—which are defined in geometric terms, unlike objects, such as bananas, whose definition includes also color. There is empirical evidence of the importance of shape. When judging category membership of newly learned objects, shape is more important than texture or size, and the bias toward shape increases with age from childhood to adulthood (Landau, Smith, & Jones, 1988). Further evidence of the importance of shape comes from experiments testing the spontaneous naming of novel objects. Again shape is seen as most important, rather than color or texture (Samuelson & Smith, 2005). Together this suggests that the entity-feature order will influence how people reason about conditionals.

The visual depiction of the sides of the die must be perceived and categorized. For each of the interpretations, the number of sides with each relevant property (relative to interpretation) must be counted. For instance for the conditional event interpretation, participants need $\lvert A \wedge B\rvert$ and $\lvert A\rvert$. There are different ways of obtaining these frequencies. One may start at the left-most die-side and count $y = \lvert A\rvert$ and then count how many of these also had the property B; denote the result x. Then the response is "x out of y." Alternatively one may begin by counting $x = \lvert A \wedge B\rvert$, store the value, and then count $y = \lvert A\rvert$, responding "x out of y." In both cases the result will be the same. At each point in the task it is possible to refresh one component, for example, the number of sides with a particular property may be recounted or the conditional statement reparsed. An additional memory component is required for goal maintenance, for example, remembering not only the conditional and counts but also the very fact that these have been remembered, what information has to be obtained next from the task presentation, and how the information must be integrated. Finally the response has to be made.

We investigated interpretations on the dice task in two experiments. We predicted that the graphical representation would reduce conjunction responses as participants are now not biased toward the joint frequencies but can count any frequencies they require. Also phrasing the task in terms of whether a conditional *holds* rather than whether it is *true* ought to reduce conjunctions. One major aim of the experiments was to determine how stable interpretations are over time. Do people reason to one interpretation and stay with that interpretation, or do they change interpretations? We also tested the effect of entity-feature order. Previous work found a response time benefit for the entity-feature order; we tested this in Experiment 2. In both experiments we also investigated whether the proportion of conditional event responses was influenced by entity-feature order. In Experiment 2, we test if the order of responses reveals the strategy used to compute probabilities. Participants who first count $\lvert A\rvert$ may wish to unburden their working memory before counting $\lvert A \wedge B\rvert$. Allowing them to do so reveals their order of processing. Moreover we predicted that response times would be faster for a conjunction rather than a conditional event interpretation, as participants need not count both $\lvert A \wedge B\rvert$ and $\lvert A\rvert$.

Experiment 1

Method

Participants. The participants were 66 students (57 women and 9 men), whose ages ranged from 20 to 40 years ($M = 23.8$; $SD = 3.5$), at the beginning of an introductory psychology course on thinking and reasoning (before conditional reasoning had been introduced) at the University of Salzburg.

Materials and procedure. For the between-participant manipulation of entity-feature order, 33 participants were assigned to the entity-feature condition, and 33 were assigned to the feature-entity condition (conditions alternated in the distribution of booklets). From the original bank of 84 items, 71 were selected such that the probability of the antecedents for both entity-feature orders were not zero. The instructions and item presentation were computer controlled and displayed on the theater screen using a data projector. Responses were given on a response sheet designed for automatic scoring (see Figure 1b). The item number was displayed on screen and on the response sheet. For the first trial, participants were given 30 s to respond. The second trial lasted 10 s, followed by a pause during which the experimenter explained that the task was about to begin. Each test trial lasted 10 s, the end of which was indicated by three beeps.

Results and Discussion

Modal responses. Responses were classified by pattern matching on numerator and denominator (i.e., assuming no simplification; see more on this later). Of the 71 items, responses to 46 could all be uniquely classified. Counting each participant's modal response type, 50 participants responded mostly with the conditional event (median 43, range 15–46), eight with the conjunction (median 27, range 17–46), and six with some other nonpredicted response (median 27, range 23–34). There was one participant responding mostly with the reversed conditional event (a score of 23) and one material conditional responder (all responses). The pattern of conditional event, conjunction, and material conditional responses matches that found previously in the literature.

Fraction simplification. Although we did not introduce the "out of" as a fraction, it is possible that some participants could have interpreted it as such and (in addition) simplified fractions. Responses were examined that could be uniquely classified even if simplification had occurred. This resulted in 1,947 responses. There was little evidence of simplification, all resulting in "other" responses when assuming no simplification, and mapping to one of the following predicted interpretations when allowing for simplification: 1.23% of the responses would map to the material conditional, 1.02% to conjunction, 0.66% to the reversed material conditional, 0.41% to the conditional event, and 0.21% to the reversed conditional event. Since this affects so few of the responses the remaining analyzes assume no simplification.

Effect of item position. As participants proceeded through the task, the proportion of conditional event interpretations increased ($r(44) = .82$, $p < .001$) and the proportion of conjunction responses decreased ($r(44) = -.73$, $p < .001$), as did reversed conditional event responses ($r(44) = -.43$, $p = .003$) and "Other" responses ($r(44) = -.36$, $p = .01$). No correlation was found between the proportion of material conditional responses and item position ($r(44) = -.01$, $p = .9$). See Figures 2a–2c for plots of the proportions of the three main interpretations as a function of item position.

One possible explanation for the convergence on the conditional event is in terms of speed-accuracy trade-off. Participants were given a limited time to respond. More time may be required to process the material using the conditional event interpretation, so those participants who appeared to shift interpretation actually had a fixed interpretation, but task adaptation allowed them to give the answer within the time available. Those who shifted from a conjunction response may first have calculated the joint probability. The absence of an effect for the entity-feature order may be because the conditional remained constant throughout the task and thus needed to be processed only once. These possibilities will be addressed in the next experiment.

Experiment 2

In this experiment we adapted the dice task for computer-controlled individual testing to (a) collect response times, (b) determine whether participants respond first with the numerator or with the denominator, (c) vary the shapes and colors in the conditionals between trials to ensure reprocessing of the conditional for each item, (d) be more careful with participant selection (to

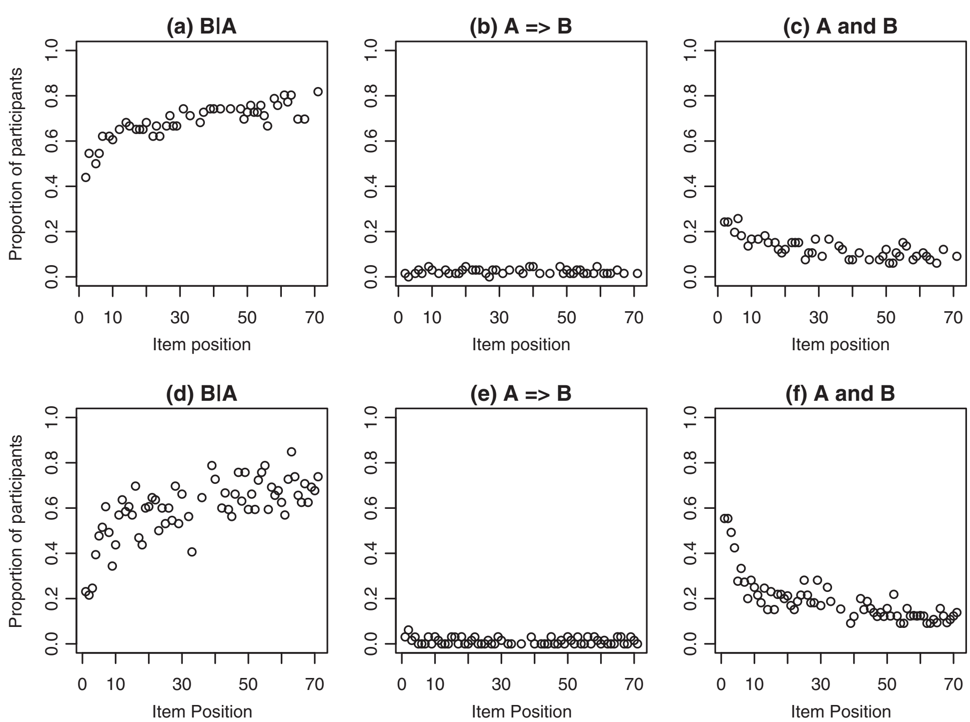


Figure 2. Proportion of participants giving a response of each class as a function of item position in Experiment 1 (a–c) and in Experiment 2 (d–f). Only the uniquely classifiable items are included.

exclude psychologists), and (e) improve experimental conditions compared to those in a lecture theater. We hypothesized that response times will be shorter for participants using a conjunction interpretation as they have to count only one joint and no marginal frequency. Further we hypothesize that if the entity is presented in the antecedent, then participants will be faster in evaluating its probability than if it is presented in the consequent.

Method

Participants. Participants were 65 students (32 women and 33 men) whose ages ranged from 18 to 30 years ($M = 22.9$; $SD = 2.9$) from the University of Salzburg, 49 of whom study a natural science, and 16 of whom study a humanities subject. Students of psychology, mathematics, or with a special background in formal logic were not included in the sample. We paid 5 Euros for participation.

Materials and procedure. A button box was designed (see Figure 1c) with a layout similar to the pen-and-paper response sheet layout used in Experiment 1. We added an extra shape (triangle) and color (green), and randomly cycled through colors and shapes to encourage participants to reprocess the conditional, thus making it more likely that an effect of entity-feature order can be detected. The areas of the shapes were adjusted so that they have the same perceivable area (Fisher & Foster, 1968). The colors were adjusted to have the same perceptual weight by fixing their chroma and luminance and varying hue (Zeileis, Hornik, & Murrell, 2009). Between-participant we crossed sex, random order (one order, forward/backward), and entity-feature order. Within-participant we varied the frequencies of shapes and colors with the constraint that the probabilities of the antecedents are not zero. Each item remained on screen until participants made their responses.

Results and Discussion

Modal responses. Counting each participant's modal response type for the 46 uniquely classifiable items, 45 participants responded mostly with the conditional event (median 40, range 19–46), 11 with the conjunction (median 42, range 20–46), two with the reversed conditional event (18 and 39), nobody with the material conditional, and seven with some other response (median 29, range 19–37). This replicates the pattern observed in Experiment 1.

Effect of item position. As participants proceeded through the task, the proportion of conditional event interpretations increased ($r(63) = .68$, $p < .001$) and the proportion of conjunction responses decreased ($r(63) = -.73$, $p < .001$), as did the proportion of reversed conditional event responses ($r(63) = -.26$, $p = .03$)[1] This replicates the convergence effect found in Experiment 1. This rules out the possibility that conjunction responses were due to people being interrupted during their computation of conditional event interpretation. There was no correlation found between item position and the proportion of material conditional responses ($r(63) = -.12$, $p = .3$). See Figures 2d–2f.

Unlike Experiment 1, no correlation was found between item position and the proportion of "other" responses ($r(63) = -.03$, $p = .8$). An explanation for this is that Experiment 1 was experimenter paced, so many participants may have taken some time at the beginning to adapt to task demands; some of these produced noise until they had adapted.

Within-participant analysis of convergence. We sought to investigate within-participants the nature of this increase in conditional event responses. Do participants smoothly increase the probability of a conditional event interpretation or is there a sudden shift in interpretation? Visual inspection of responses suggested that many participants shifted suddenly to a particular interpretation after some time. Thus we decided to investigate interpretation shifts systematically to detect for whom and when this occurred. To find a shift point for each participant, we used the following simple algorithm:

1. Let $S = \langle s_1, \ldots, s_{71} \rangle$ denote the binary sequence of 71 conditional event scores. $C = \langle c_1, \ldots, c_{71} \rangle$ denotes a sequence of 71 scores, where each element of C represents how many different interpretations a 1 in the conditional event score could represent, e.g., if the ith response could be either a conditional event or conjunction, then $s_i = 1$ and $c_i = 2$. For a given i, $c_i \in [0, 5]$ (0 if the response is "other").

2. Use these two sequences to create a weighted sequence, $W = \langle w_1, \ldots, w_{71} \rangle$: if $c_i = 0$, then set $w_i := 0$, as this response is "other;" otherwise set $w_i := s_i/c_i$.

3. For every $i \in [2, 71]$, compute the proportions $l_i = \sum_{j=1}^{i-1} w_j/(i-1)$ and $r_i = \sum_{j=1}^{71} w_j/(71 - i + 1)$. Note that r_i includes position i.

4. The change point is found by maximizing $r_i - l_i$. When there is more than one i where this difference is maximal, we take the first.

We also computed the modal interpretation before and after this change point and the proportion of responses of these modal types, using the 46 uniquely classifiable responses. Just over half of the participants (36, around 55%) shifted from some other interpretation to the conditional event interpretation. Of these, the majority (29, around 80%) shifted from the conjunction interpretation, three from the reversed conditional event ($A|B$), three from some nonclassifiable response, and only one from the material conditional—but in its reversed form ($B \supset A$).

The earliest shift occurred at Item Position 2 (one participant), with 64% of those who shift doing so at least by Position 8. Figure 3a shows the distribution of the change points. Figures 3b and 3c show the proportion of responses of the modal type before and conditional event after (and including) the change point. As may be seen, most participants are very consistent once they have shifted to the conditional event (mean proportion of conditional event responses after the shift is .93, $SD = .1$).

A further Bayesian change point analysis for binary data was performed (Tan, Tian, & Ng, 2010). Only the 46 uniquely classifiable responses were analyzed. The sequence was coded 1 for the conditional event interpretation and 0 otherwise. A single–change point model was chosen with a uniform prior distribution on putative positions (45 putative change points,

[1] Correlations were computed using the original item positions (1 to 71), not their relative position (1 to 46). Because data from the two random orders were pooled, 65 rather than 46 pairs of values resulted, as data were available for a particular item position in only one direction for 19 positions.

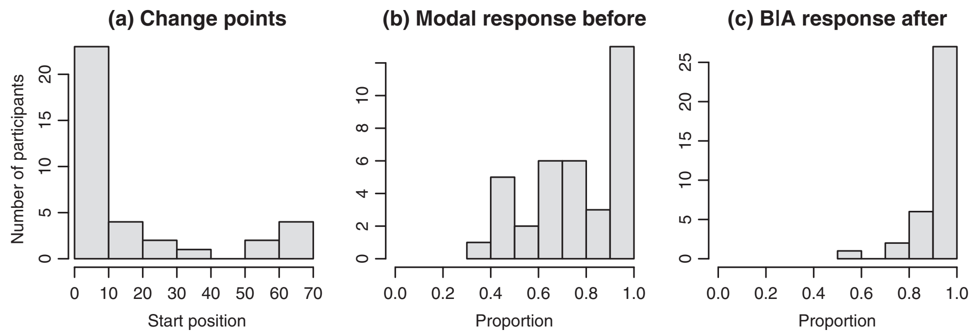


Figure 3. (a) Distribution of change points for the 36 participants who shifted. (b) Proportion of responses before the change point that are of the modal class. (c) Proportion of responses after (and including) the change point that are consistent with the conditional event interpretation.

giving a constant probability of $1/45 = .022$ at each position). For each participant, the posterior probability distribution of change points was computed. The mode of each posterior distribution was used to determine the change point. The change points computed by the two methods for the 36 people who shift to the conditional event (as determined by the method above) were highly correlated ($r(34) = .91$, $p < .001$), supporting the analysis. The modes of the posterior distributions had a mean of .6 ($SD = .2$), well away from the uniform distribution.

We also have some self-report data from the participants on their strategies. Participant 34 (who settled into a conjunction interpretation) said: "I only looked at the shape and the color, and then always out of 6; this was the quickest way." Participant 37, who shifted from the conjunction to the conditional event, said: "In the beginning [I] always [responded] 'out of 6,' but then somewhere in the middle . . . Ah! It clicked and I got it. I was angry with myself that I was so stupid before." Five participants spontaneously reported when they shifted during the task, for example, saying, "Ah, this is how it works." Such unprompted comments are typical indicators of insight effects (Bowden, Jung-Beeman, Fleck, & Kounios, 2005).

Response production order. Fourteen participants (around 20%) pressed a button from the bottom row first at least once. Eight of these did so exactly once, and the remainder between 10–40 times out of 71 responses. Only one conjunction response from one participant was made by pushing the bottom button first. For conditional event responses, only four participants pressed the bottom button first a nonnegligible number of times: 15–25. Thus the hypothesized benefit of unburdening working memory by first responding with the denominator for conditional event responses has not received strong support.

Response time analysis. The mean response time for the numerator was 6163 ms ($SD = 2079$ ms) and for the denominator was 985 ms ($SD = 580$ ms).

We tested our hypothesis that participants would be faster for a conjunction versus a conditional event response using mixed-effects models. Models were fitted using the lme4 package (Baayen, Davidson, & Bates, 2008; Bates, Maechler, & Dai, 2008) in R (www.r-project.org). HPD intervals were estimated using MCMC draws from the posterior distributions. Log-likelihood ratio tests, and Akaike's information criterion (AIC), derived from the maximum log-likelihood estimates and penalized for the number of parameters, were used to compare fitted models. The basic model structure was as follows:

$$\begin{aligned} log(RT_{ip}) \; = \; & \beta_0 + \gamma_{0p} + \gamma_{1i} + \beta_1 \cdot pos_{ip} + \beta_2 \cdot pos_{ip}^2 \\ & + \beta_3 \cdot [A|B]_{ip} + \beta_4 \cdot [A \wedge B]_{\mathrm{ip}} + \beta_5 \cdot [A \supset B]_{ip} \\ & + \beta_6 \cdot [B \supset A]_{ip} \; + \beta_7 \cdot Other_{ip} + \varepsilon_{ip} \end{aligned}$$

where p is a participant, i an item, and pos is the item position (added with a quadratic term to model the overall speedup of responses). The terms $[A|B]_{\mathrm{ip}}$, $[A \wedge B]_{ip}$, $[A \supset B]_{ip}$, $[B \supset A]_{ip}$, and $Other_{ip}$ were coded 1 if the response was according to the corresponding classification and 0 otherwise. The conditional event, $B|$A, interpretation is the baseline category against which all others were compared, thus does not appear as a predictor. The coefficient γ_p represents between-participant variation in mean response time and γ_i represents participant-invariant effects of items.

First the effect of the response type was tested. Adding this variable improved the fit of the model (ΔAIC $= -8$, log-likelihood ratio (LLR) $\chi^2(5) = 18.1$, $p = .003$). As predicted, conjunction responses were faster than the conditional event (95% HPD interval $\in [-0.15, -0.04]$). The mean difference predicted using the model's fixed effect terms was 503 ms. HPD intervals for all other interpretation-type predictors versus the conditional event included 0. There was no main effect of entity-feature order, however participants were slower when giving a conjunction response in the feature-entity condition than in the entity-feature condition (95% HPD interval $\in [0.06, 0.26]$).

Effect of entity-feature order on interpretation. Figure 4 shows the proportion of participants giving a conditional event or conjunction response as a function of item position, split by entity-feature order. A generalized linear mixed effect model was fitted with binomial errors and a *logit* link. The dependent variable was the probability of a conditional event response. As before predictors were added for item position (pos). Also a predictor, *order*, was added for the entity-feature order: 1 if feature-entity and 0 if entity-feature. We found no main effect of entity-feature order (ΔAIC $= 2$, LLR $\chi^2(1) = 0.03$, $p = .9$), however there was an interaction between item position and entity-feature order (ΔAIC $= -17$, LLR $\chi^2(1) = 18.5$, $p < .001$). The final model chosen was as follows:

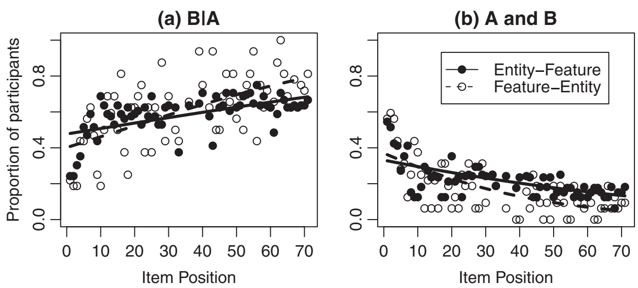


Figure 4. Proportion of participants giving a response of each class as a function of item position split by feature-entity order. Only the uniquely classifiable items are included.

$$logit(P(y_{ip} = 1)) = \beta_0 + \gamma_{0p} + \gamma_{1i} + \beta_1 \cdot order_{ip} + \beta_2 \cdot pos_{ip} + \beta_3 \cdot pos_{ip} \cdot order_{ip} + \varepsilon_{ip}$$

Figure 5 shows predictions from the model's fixed-effect estimates. At the beginning of the task, participants in the entity-feature condition were more likely to use a conditional event interpretation than in the feature-entity condition, but then at the end of the task the reverse was found: conditional events were more common in the feature entity condition than in the entity feature condition.

General Discussion

We set out to investigate (a) what are the dominant interpretations of uncertain conditionals on our dice task, (b) how do response times relate to interpretation, (c) how stable is interpretation over time, and (d) how does entity-feature order in the conditional influence interpretation?

Participants' Individual Item and Modal Responses

Replicating previous results, the conditional event was the most common interpretation of the if-then (modal response for 76% of participants in Experiment 1 and 69% in Experiment 2), followed by the conjunction (12% in Experiment 1 and 24% in Experiment 2). Material conditional responses were negligible. Conditional event interpretations are more common than found previously in some of the literature, for example, just over half in the experiment by Evans, Handley, and Over (2003). Gauffroy and Barrouillet (2009) independently developed a graphical test with the same out-of response format we used. They found similar patterns of responses as did we: slightly fewer conditional event responses (a little under 60%) and more conjunctions (30%).

One motivation for presenting the joint frequencies graphically was so as not to bias people toward the conjunction interpretation. Participants selected and obtained by counting the frequencies they required for their interpretation, for instance $|A|$ could be counted directly and it was unnecessary to compute it from $|A \wedge B| + |A \wedge \neg B|$, as is the case for the standard probabilistic truth table tasks where the joint frequencies are provided numerically. Conjunction probabilities did not disappear, ruling out the possibility that they were due only to simple matching on the numerical frequencies provided by previous tasks. However, binding processes in the visual system could still bias people toward the joint frequencies, that is, the initial representation of the sides of a die is in terms of undifferentiated wholes, for example, red squares and blue circles, rather collections of squares, circles, red objects, and blue objects. This is consistent with evidence that integrated object representations, rather than representations of individual features, are constructed and held in visual working memory (e.g., Gajewski & Brockmole, 2006; Luck & Vogel, 1997). As explained in the introduction, another possibility is that conjunction responses resulted from an interpretation mapping the natural language if-then to the conditional event but with a *task* interpretation of inferring the probability that the conditional event is *true* (Edgington, 2003). We asked the question, "How sure can you be that the following conditional holds?" rather than referring to *true* in order to make this possibility less likely.

Our experiments provide more decisive evidence than previous experiments against the material conditional and in favor of the conditional event interpretation. The task used in previous experiments (e.g., Evans, Handley, & Over, 2003; Oberauer et al., 2007; Oberauer & Wilhelm, 2003) necessitated an indirect scoring of interpretation. For instance, it was assumed by Evans, Handley, & Over (2003) that if people gave the material conditional interpretation for "if A, then B," then they would also do so for the contrapositive, "if not- B, then not- A," as $A \supset B$ is equivalent to $\neg B \supset \neg A$. This equivalence does not hold for the conditional event interpretation, so a low correlation in probabilities produced for these two conditionals (and fixed frequency) was interpreted as evidence against the material conditional interpretation and in favor of the conditional event. This assumption is problematic. It is possible that participants would reason to a material conditional interpretation but find the contrapositive more difficult to reason about than the direct conditional, resulting in different probabilities. The material conditional probability should correlate positively with the frequency of false antecedent cases. The opposite effect is interpreted as ruling out the material conditional. Again this is a very indirect way to infer the interpretation and depends on relationships between item responses rather than allowing the interpretation of each individual item to be inferred. Oberauer and Wilhelm (2003) used two indexes to infer interpretations. One was computed from the ratio of the frequency of $A \wedge B$ cases to the frequency of $A \wedge \neg B$ cases and the other from the frequency of $A \wedge B$ cases. A correlation of responses with the ratio index indi-

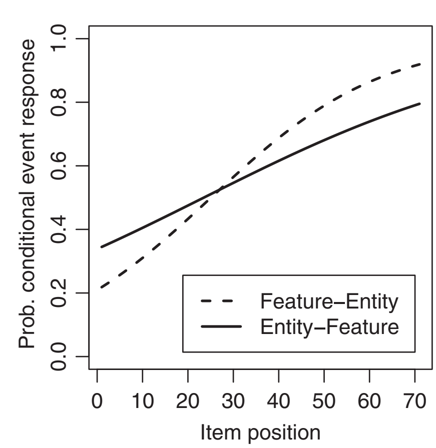


Figure 5. Interaction between item position and entity-feature order. Prob. = Probability of.

cated a conditional event interpretation and, with the conjunction index, a conjunction interpretation. However, the task we used enabled direct comparison of competence model predictions with the response people gave for each item. A further improvement of our task over other tasks is the systematic enumeration of all possible patterns on two-dimensions painted on to a six-sided die, thus providing strong evidence that the conditional event interpretation is not limited to a subset of frequency combinations.

By separating the numerator and denominator responses and recording which was input first, we are able to provide constraints on the preferred ordering of processing of probabilities. For the vast majority of participants, the numerator was entered first, followed by the denominator. Only four participants gave conditional event responses by first entering the denominator. This is evidence against our hypothesis that participants first count and respond with the number of sides of the die where the antecedent—which comes first in the conditional—is true (equal to the denominator), and then following this count and respond with the cardinality of the subset of those sides where the consequent is true. It seems, on the contrary, that participants first count conjunction cases, and then consider how many of all the cases are relevant. This is consistent with the first stage of the psychological implementation of the Ramsey test suggested by Evans, Handley, and Over (2003) where $P(A \wedge B)$ is compared with $P(A \wedge \neg B)$. For the dice task, it is more likely that the second stage is to compute $P(A)$ rather than $P(A \wedge \neg B)$. For the numerically presented tasks, it is easy to see why the model suggested by Evans, Handley, & Over might be more appropriate. Consider the example from earlier:

1 yellow circle

4 yellow diamonds

16 red circles

16 red diamonds

The conditional is "If the card is yellow, then it has a circle printed on it." It is plausible that participants make a comparison of the two joint frequencies, $|\text{yellow} \wedge \text{circle}|$ and $|\text{yellow} \wedge \neg\text{circle}|$, as a way to avoid explicit arithmetic, especially because the response format required answers on a one-dimensional bounded scale. Evans, Handley, and Over (2003) used a discrete 5-point scale, which biases participants toward a ranked judgment. Oberauer et al. (2007) asked participants to respond on a scale between 0 and 100 and not to use explicit calculation: "Don't calculate the probability. Give a subjective estimation!" (p. 795). To give an exact response, participants must first infer how to extract the denominator from two joints and then, in an additional step rescale, for example, by computing $1/(1 + 4) \cdot 100 = 20$, so again, for many problems, a qualitative comparison process is more likely to be used. We have provided firm evidence that not only are people able to estimate probabilities of conditionals but they can also calculate point probabilities, for both conjunction and conditional event interpretations.

Conditional event responses took more time to process than did conjunction responses. One way to explain the results is through dual-process theories (see Evans, 2008, for a review). These postulate the existence of at least two types of process. Type 1 processes are fast, operate in parallel outside of conscious awareness, and do not require working memory. Type 2 processes are slower, serial, operate with an associated conscious awareness of their operation, and impose a working memory burden. Evans (2009) introduced the idea of Type 3 processes, which are hypothesized to deal with conflict resolution between Type 1 and Type 2 processes. Type 1 processes could then be responsible for conjunction responses and Type 2 for conditional event responses (this is also consistent with Barrouillet, Gauffroy, & Lecas, 2008a, see later). Compatible with this suggestion is the result by Evans et al. (2007) that those who give conditional event responses tend to have a higher score on the AH4 intelligence test, interpreted as reflecting greater working memory and thus higher likelihood of engaging Type 2 processes, than do those who give conjunction responses. Another explanation is that a deliberate Type 2 process is required to ignore the false antecedent cases.

The slower conditional event responses are also predicted by an information analysis of how long it would take to obtain the appropriate frequencies for conditional event versus conjunction interpretations. For the conjunction response, the denominator is always 6, whereas for the conditional event response the denominator varies between trials. Our task analysis is suggestive of some of the details of processes that must be required to solve the task. Processes for recognizing composite bound visual objects—especially simple objects like squares and circles—are likely to be of Type 1. The processes required for counting and remembering how many sides of the die have a particular property are more likely to require Type 2 processing. By this account, both conjunction and conditional event responses depend on Type 2 processes; however, for the conjunction, there is less of a burden on Type 2 processes as less working memory is required to store only the conditional and numerator and not, in addition, the denominator.

Shifts of Interpretation

One question often neglected in the psychology of reasoning is how stable interpretations and strategies of inference are within-participant. We found evidence of interpretation shifts. For instance in Experiment 2, 55% of participants shifted to a conditional event response during the task, and 80% of these shifted from conjunction responses. We also found evidence that the shift to the conditional event interpretation was later for feature-entity order compared to the entity-feature order and that, at the end of the task, conditional event interpretations became more common for the feature-entity than entity-feature order.

People must *interpret* the task they are supposed to solve before reasoning from that interpretation, that is, *derivation*, can proceed (Stenning & van Lambalgen, 2008). It is difficult to distinguish between effects due to individual differences in interpretation and those due to differences in derivation. An "error" with respect to a particular competence model may be viewed instead as evidence of a different interpretation. In Experiment 1, but not in Experiment 2, the proportion of "other" responses decreased as a function of item position. Such "other" responses are likely to represent noise, which we interpret as a failure of derivation because Experiment 1 was experimenter-paced. In Experiment 2, which was participant-paced, the correlation between proportion of "other" responses and item position disappears. The most common shift in

Experiment 2 was from a conjunction response to a conditional event response. This is indicative of a shift in interpretation rather than an improvement in derivation strategy.

Changes of interpretation have been observed in an experiment using a nonprobabilistic truth table task (Politzer, 1981), though the item-by-item progression of interpretation was not investigated. This effect was argued to be cued by the process of going through the truth table cases. It could be that processing many different dice has an analogous effect, though it is not clear what processes enable the effect. Related work is that of learning rules to solve Raven's Progressive Matrices (Verguts & Boeck, 2002). Items that required the same rules for solution (according to a rule-based model) were repeatedly presented to investigate transfer and learning. The authors argue that participants become more fluent in their rule use over repeated applications. However they discovered only an effect of rule learning when feedback was given on response accuracy. Self-discovered rules tended not to be as fluently reapplied to later items. In our task, no feedback was given, yet participants shifted to a conditional event interpretation.

Insight is defined as the effect of suddenly understanding how to solve a problem after a period of impasse, often accompanied with an "'Aha!" feeling (Bowden et al., 2005). Our results suggest that participants who shifted interpretation demonstrated such an effect, both by qualitative shifts in response type and also (for some participants) by spontaneous self-reports of insight. Problems used to study insight, for example, anagrams, usually have a clear goal; the difficulty comes from how to achieve that goal from the starting state. For our reasoning task, participants suffer no impasse: they begin by computing the conjunction probability. The interpretation shift is thus a shift in the representation of the goal rather than how to achieve the goal. As discussed in the task analysis, it is possible that participants suffer a failure of goal memory. This could result in the conjunction responses. This is plausible because the conditional event probability depends on the conjunction probability. However it would have to be a very systematic failure of goal memory: When the experiment is self-paced, participants do not produce noise before shifting to the conditional event. As discussed above, it is also possible that participants are biased toward the conjunction probabilities because of visual binding processes and task adaptation takes time to overcome this. Processes computing conditional event and conjunction responses have competing goals: The processes resulting in a conditional event do not win the competition until after some items have been processed in those who shift. Although the shift in responses is sudden, it is still possible that Type 1 parallel processes, operating outside of participants' awareness, incrementally compute two (or many more) interpretations then, after some time, the most likely interpretation is inferred to be the conditional event. A similar incremental account has been given of sudden "pop-out" solutions in anagram solving (Novick & Sherman, 2003).

The account of mental models by Johnson-Laird and Byrne (2002) posits that there are two main kinds of mental representation for conditionals that are as independent as possible from context and background knowledge, "if *A*, then *B*." One consists of a single explicit mental model that represents the $A \wedge B$ case plus an implicit model representing that other possibilities exist. If participants build this representation but forget the implicit model, then a conjunction response results. If they remember the implicit model, then (according to the account) some naive participants give a response similar to what we would call a conditional event response. The other representation consists only of explicit models, which, in addition to the $A \wedge B$ case, also represents the cases $\neg A \wedge B$ and $\neg A \wedge \neg B$. The representation with these explicit models gives the material conditional interpretation of natural language conditionals. Mental models theory predicts a shift from conjunction to material conditional interpretations, which we do not find.

Girotto and Johnson-Laird (2004) present verbal protocol evidence (see also Byrne & Johnson-Laird, 2010) that many participants rephrase conditionals when reasoning about them. So, for instance, the task, "What is the probability that if the side shows a square, then the side shows red?" would be rephrased as "If the side shows a square, then what is the probability that the side shows red?" This is a *narrow-scope* interpretation of the probability function, which reduces to the modal fallacy (see Over, Hadjichristidis, Evans, Handley, & Sloman, 2007). Our task avoids a narrow-scope reading and emphasizes a *wide-scope* interpretation by placing the conditional in a box, and by phrasing the task as follows:

How sure can you be that the following sentence holds?

If the side shows a square, then the side shows red

To explain the within-participant shift of interpretation from conjunction to conditional event according to the account by Girotto and Johnson-Laird (2004), first the task would have to be interpreted as referring to the whole conditional, without rephrasing, and a single model constructed to represent the case *square* $\wedge$ *red* and no implicit model. Then, when the shift occurred, the task would have to be represented with the rephrased conditional and attention restricted to those cases where the side shows a square. The task is then to infer how sure you can be that the side shows red. The result is indistinguishable from the Ramsey test. It is also not clear why a process leading to rephrasing would initially leave the correct interpretation of the conditional (according to Girotto & Johnson-Laird) and then, given more trials, subsequently give an incorrect representation.

The within-participant shifts of interpretation found in our student sample follow the trend shown across participants in developmental studies of conditional interpretation (Barrouillet, Gauffroy, & Lecas, 2008a). Just over half of children aged around 9 years give a conjunction interpretation and conditional event interpretations are entirely absent. Conditional event interpretations appear at around 15 years (around 30% of participants), and then, by the late 20s, the conditional event is the most common interpretation (55% of participants). This pattern of development has also been found on a probabilistic truth table task (Gauffroy & Barrouillet, 2009), though shifted by 3 years: 12-year-olds mostly responded with a conjunction (at around 80%); at age 16 years, conditional event began to appear (25% if participants); finally, in undergraduate students, the proportion of conjunctions remained around the same and the conditional event rose to 55% of participants, as commonly found in the literature.

Barrouillet, Gauffroy, & Lecas (2008a) revised mental models theory to accommodate the results (see also the exchange Oberauer & Oaksford, 2008b; Barrouillet, Gauffroy, & Lecas, 2008b; Oberauer & Oaksford, 2008a). Following Evans (2006), their theory posits two types of processes: heuristic and analytic. Heuristic processes are seen as independent of working memory capacity and general intelligence. According to the account, heuristic processes generate conjunction interpretations at all levels of development. The conditional event interpretation is hypothesized to require two additional mental models to represent the possibilities where it is void, so development of interpretation is explained by development of working memory capacity.

We agree that a Type 1 process focuses attention on the conjunction cases. For instance for the conditional, "if the side shows a square, then the side shows red," a conjunction response results from counting only sides mentioned in the conditional, red squares. However we do not think it is necessary to postulate the representation suggested by Barrouillet, Gauffroy, & Lecas (2008a). Each case can be assigned the relevant semantic value (*true*, *false*, or *void*) and all possibilities need not be stored in working memory to do so because possibilities are provided visually in the task. An alternative explanation is that initially all possibilities are relevant to participants because initially they do not know what those possibilities are. In order to reason to the conditional event interpretation, cases for which the antecedent is false need to be selectively ignored, a process requiring inhibitory processing. This is not to say that all six cases are represented in working memory, although it is likely that the number of cases is. In addition to working memory capacity increasing through development, inhibitory processes also improve (e.g., Huizinga, Dolan, & van der Molen, 2006), so this account is compatible with the developmental trend. In adults who shift interpretation, it could be that inhibitory processes take some time to engage, for instance as a function of time to adapt to the overall task demands. This account is also compatible with the idea of a two-stage Ramsey test introduced earlier (Evans, Handley, & Over, 2003), except that in those who switch, it is not a failure of working memory which stops the second stage but rather a failure of inhibitory processes to allow the second stage to be computed.

Developmentally, between conjunction and conditional event interpretations, a "defective biconditional," or what is better named a "biconditional event", interpretation appears (Barrouillet, Gauffroy, & Lecas, 2008a; Gauffroy & Barrouillet, 2009). This interpretation is equivalent to $A \wedge B | A \vee B$. The developmental trend can be expressed as a narrowing of the hypothetical scope. For the conjunction interpretation, the whole event space is seen as relevant. The biconditional event reduces the relevant events to those cases where either A or B is true. Then finally the conditional event interpretation focuses on cases where A is true. Again, an inhibitory account of this process is compatible with the trend, with increasing ability to narrow attention on the appropriate subset of the event space developing with age.

Although some researchers argue that a homogeneous inhibition construct can be measured using latent variable modeling (e.g., Friedman et al., 2008; Miyake et al., 2000), there is much evidence of heterogeneity in tasks purporting to measure inhibition, for instance low correlations between the measures. If inhibition does explain the shifts, then the exact inhibitory processes required for probabilistic tasks still need to be investigated. Also, inhibition cannot be the whole story. Those who reason to a conditional event interpretation of the task have to have appropriate knowledge once the initial conjunctive interpretation is inhibited. The same is true of accounts based on working memory capacity. Capacity to build representations is necessary but not sufficient to reason to the conditional event interpretation.

A number of results show that biases observed in the responses given by children can still be detected in the responses given by adults. For instance 3-year-old children have difficultly on dimensional card sorting tasks where one is shown cards with pictures on two dimensions and the task is first to sort them by shape and then by color (Frye, Zelazo, & Palfai, 1995). Children have difficultly switching between dimensions. Diamond and Kirkham (2005) found that, although adults were successfully able to switch the dimensions to be sorted, there was a lingering effect of lengthened response times postswitch. Children around 3 years of age have problems on false belief tasks where one has to inhibit what one knows to be the case, and reason instead from the perspective of an actor whose belief differs from reality (Wimmer & Perner, 1983). Birch and Bloom (2007) found that when adults are asked for the probability the actor will search at various locations for an object, they also tend to show traces of their own true belief about where the object is hidden rather than the false belief of the character from whose perspective they are supposed to be reasoning. We speculate that the present results suggest that uncertain conditional interpretation is another area where, at least initially, cognitive processes developing early strongly determine the responses adult participants give.

We found that the convergence on a conditional event interpretation was moderated by the entity-feature order: At the beginning of the task, participants were more likely to give a conjunction interpretation and less likely to give a conditional event interpretation, for the feature-entity order (e.g., "if the side shows red, then it shows a square") than for the entity-feature order (e.g., ""if the side shows a square, then it shows red"). By the end of the task, the opposite was the case: Conditional event interpretations were more likely for the feature-entity order than the entity-feature order. We also found that conjunction responses were slower in the feature-entity condition than in the entity-feature condition. To the best of our knowledge, current versions of mental models theory predict no difference in the inferences drawn between the entity-feature and feature-entity order, as the types of the entries (i.e., whether they are shapes or colors) are abstracted away in the representation.

How then are we to explain the moderating effect of entity-feature order? Kleiter (1986) found that response times in verifying verbal analogies were faster when the entity was first and feature second. The explanation for this was that it is easier to attach a feature to an already existing representation of a concrete entity (e.g., an eye) than vice versa. The task used by Kleiter had no explicit visual component, and the relationships between entities and features were retrieved from semantic memory. In our task, the entities and features referred to by antecedent and consequent (e.g., *square* and *red*) must also have been retrieved from semantic memory, otherwise the objects could not have been identified and counted; however, in semantic memory, the shapes and colors were likely to have been independent. It is participants' recognition of the objects in the graphical presentation of sides of the dice that induces the dependency between entity and feature, so it is likely that a representation of the conditional will involve both semantic representations and visual elements in working memory. There is evidence that memory for colors is better than memory for shapes on visuospatial working memory tasks (Delvenne, Cleeremans, & Laloyaux, 2010; Song & Jiang, 2006), suggesting that representations of shapes are more complex than for colors. Conditional event interpretations could be more common by the end of the task for the feature-entity (i.e., color-first) ordering because the reduced feature complexity imposes less of a burden on inhibitory function.

Conclusion

In two experiments, the percentage of participants interpreting the natural language conditional as a conditional event increased from around 40% at the beginning of the task to nearly 80% by the end. Most shifted from a conjunction interpretation. This interpretation shift was moderated by the entity-feature order of terms in the conditionals. Replicating previous results, we found that the conditional event was the most common interpretation of the natural language *if*. Finally, conditional event responses took longer to produce than did conjunction responses. Not only do different people reason to different interpretations but individuals shift interpretations during a task, and these shifts are affected by the content of the material to be reasoned about. Studying trajectories of interpretation change reveals participants' inferences about the correctness of their interpretation. The interpretation by the end of the task for the majority of participants corresponds with that proposed by a long tradition in philosophy for the meaning of *if*. This is in contrast to the core meaning of conditionals proposed by mental models theory. Future work is needed to clarify when and for whom these shifts of interpretation occur and what cues can facilitate or impede the process.

References

Adams, E. W. (1975). *The logic of conditionals.* Dordrecht, the Netherlands: Reidel.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390–412.

Barrouillet, P., Gauffroy, C., & Lecas, J.-F. (2008a). Mental models and the suppositional account of conditionals. *Psychological Review, 115,* 760–771. Available from http://dx.doi.org/10.1037/0033-295X.115.3.760

Barrouillet, P., Gauffroy, C., & Lecas, J.-F. (2008b). Postscript: A good psychological theory of reasoning must predict behavior and explain the data. *Psychological Review, 115,* 771–772.

Bates, D., Maechler, M., & Dai, B. (2008). lme4: Linear mixed-effects models using s4 classes [Computer software manual]. Available from http://lme4.r-forge.r-project.org/ (R package Version 0.999375–28)

Bennett, J. (2003). *A philosophical guide to conditionals.* Oxford: Oxford University Press.

Birch, S. A. J., & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychological Science, 18,* 382–386. Available from http://dx.doi.org/10.1111/j.1467-9280.2007.01909.x

Bowden, E., Jung-Beeman, M., Fleck, J., & Kounios, J. (2005). New approaches to demystifying insight. *Trends in Cognitive Sciences, 9,* 322–328.

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. (Eds.). (1998). *Mental logic.* Mahwah, NJ: Lawrence Erlbaum Associates.

Byrne, R. M. J. (1989). Suppressing valid inferences with conditionals. *Cognition, 31,* 61–83.

Byrne, R. M. J., & Johnson-Laird, P. N. (2010). Models redux: Response to Evans and Over. *Trends in cognitive sciences, 14,* 6.

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

Da Silva Neves, R., Bonnefon, J.-F., & Raufaste, E. (2002). An empirical test of patterns for nonmonotonic inference. *Annals of Mathematics and Artificial Intelligence, 34*(1–3), 107–130.

de Finetti, B. (1937/1980). Foresight: Its logical laws, its subjective sources. In J. H. E. Kyburg & H. E. Smokier (Eds.), *Studies in subjective probability* (pp. 55–118). New York: Robert E. Krieger Publishing Company. (English translation by H. E. Kyberg, Jr.)

Delvenne, J.-F., Cleeremans, A., & Laloyaux, C. (2010). Feature bindings are maintained in visual short-term memory without sustained focused attention. *Experimental Psychology, 57,* 108–116.

Diamond, A., & Kirkham, N. (2005). Not quite as grown-up as we like to think. *Psychological Science, 16,* 291–297.

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. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review, 13,* 378–395.

Evans, J. St. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59,* 255–278.

Evans, J. St. B. T. (2009). How many dual-process theories do we need? One, two, or many? In J. St. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 33–54). Oxford University Press.

Evans, J. St. B. T., Handley, S. J., Neilens, H., & Over, D. E. (2007). Thinking about conditionals: A study of individual differences. *Memory and Cognition, 35,* 1772–1784.

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

Evans, J. St. B. T., Over, D. E., & Handley, S. J. (2005). Suppositions, extensionality, and conditionals: A critique of the mental model theory of Johnson-Laird and Byrne (2002). *Psychological Review, 112,* 1040–1052.

Fisher, G., & Foster, J. (1968). Apparent sizes of different shapes and the facility with which they can be identified. *Nature, 219*(5154), 653–654.

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General, 137,* 201–225.

Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development, 10,* 483–527.

Gajewski, D. A., & Brockmole, J. R. (2006). Feature bindings endure without attention: Evidence from an explicit recall task. *Psychonomic Bulletin & Review, 13,* 581–587.

Gauffroy, C., & Barrouillet, P. (2009). Heuristic and analytic processes in mental models for conditionals: An integrative developmental theory. *Developmental Review, 29*(4), 249–282. Available from http://dx.doi.org/10.1016/j.dr.2009.09.002

Gilio, A. (2002). Probabilistic reasoning under coherence in System P. *Annals of Mathematics and Artificial Intelligence, 34,* 5–34.

Girotto, V., & Johnson-Laird, P. (2004). The probability of conditionals. *Psychologia, 47,* 207–225.

Huizinga, M., Dolan, C. V., & van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia, 44,* 2017–2036.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness.* Cambridge, England: Cambridge University Press.

Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review, 109,* 646–677.

Johnson-Laird, P. N., Byrne, R. M. J., & Girotto, V. (2009). The mental model theory of conditionals: A reply to Guy Politzer. *Topoi, 28,* 75–80.

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

Johnson-Laird, P. N., & Tagart, J. (1969). How implication is understood. *The American Journal of Psychology, 82,* 367–373. Available from http://www.jstor.org/stable/1420752

Kleiter, G. D. (1986). Solving analogies by building propositions. In F. Klix & H. Hagendorf (Eds.), *Human memory and cognitive capabilities* (pp. 977–986). Amsterdam: Elsevier.

Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development, 3,* 299–321.

Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making, 21,* 37–44.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*(6657), 279–280.

Manktelow, K., Over, D. E., & Elqayam, S. (2011). Paradigms shift: Jonathan Evans and the science of reason. In K. Manktelow, D. E. Over, & S. Elqayam (Eds.), *The science of reason: A Festschrift for Jonathan St. B. T. Evans* (pp. 1-–16). Hove, England: Psychology Press.

Marr, D. (1982). Vision: *A computational investigation into the human representation and processing of visual information.* New York: W. H. Freeman and Company.

Miyake, A., Friedman, N. P., Emerson, M., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex 'frontal lobe' tasks: A latent variable analysis. *Cognitive Psychology, 41,* 49–100.

Novick, L. R., & Sherman, S. J. (2003). On the nature of insight solutions: Evidence from skill differences in anagram solution. *The Quarterly Journal of Experimental Psychology, A, 56,* 351–382.

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

Oaksford, M., & Chater, N. (2009). Précis of Bayesian rationality: The probabilistic approach to human reasoning. *Behavioral and Brain Sciences, 32,* 69–120.

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., & Oaksford, M. (2008a). Postscript: Still in search of a good theory of reasoning—Rejoinder to Barrouillet, Gauffroy, and Lecas (2008). *Psychological Review, 115,* 778. Available from http://dx.doi.org/10.1037/0033-295X.115.3.778

Oberauer, K., & Oaksford, M. (2008b). What must a psychological theory of reasoning explain? Comment on Barrouillet, Gauffroy, and Lecas (2008). *Psychological Review, 115,* 773–778. Available from http://dx.doi.org/10.1037/0033-295X.115.3.773

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

Over, D. E. (2009). New paradigm psychology of reasoning. *Thinking & Reasoning, 15,* 431–438.

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.

Pfeifer, N., & Kleiter, G. D. (2005a). Coherence and nonmonotonicity in human reasoning. *Synthese, 146*(1–2), 93–109.

Pfeifer, N., & Kleiter, G. D. (2005b). Towards a mental probability logic. *Psychologica Belgica, 45,* 71–99.

Pfeifer, N., & Kleiter, G. D. (2006). Inference in conditional probability logic. *Kybernetika, 42,* 391–404.

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

Pfeifer, N., & Kleiter, G. D. (2010). The conditional in mental probability logic. In M. Oaksford & N. Chater (Eds.), *Cognition and conditionals: Probability and logic in human thoughts* (pp. 153–173). Oxford, England, and New York: Oxford University Press.

Pfeifer, N., & Kleiter, G. D. (2011). Uncertain deductive reasoning. In K. Manktelow, D. E. Over, & S. Elqayam (Eds.), *The science of reason: A Festschrift for Jonathan St. B. T. Evans* (pp. 145—166). Hove, England: Psychology Press.

Politzer, G. (1981). Differences in interpretation of implication. *The American Journal of Psychology, 94,* 461–477. Available from http://www.jstor.org/stable/1422257

Politzer, G. (2007). Reasoning with conditionals. *Topoi, 26,* 79–95.

Pygame Community. (2008). Pygame (Version 1.8.1) [Computer software]. Retrieved from http://www.pygame.org/

Python Software Foundation.. (2008). Python (Version 2.6.1) [Computer software]. Retrieved from http://www.python.org

Ramsey, F. P. (1929/1990). General propositions and causality. In D. H. Mellor (Ed.), *Philosophical papers* (pp. 145–163). Cambridge, England: Cambridge University Press. (A scan of the original handwritten manuscript is available at http://www.dspace.cam.ac.uk/handle/1810/194722)

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

Samuelson, L. K., & Smith, L. B. (2005). They call it like they see it: Spontaneous naming and attention to shape. *Developmental Science, 8,* 182–198.

Schroyens, W., Schaeken, W., & Dieussaert, K. (2008). 'The' interpretation(s) of conditionals. *Experimental Psychology, 55*(3), 173–181.

Song, J. H., & Jiang, Y. (2006). Visual working memory for simple and complex features: An fMRI study. *Neuroimage, 30,* 963–972.

Stenning, K., & van Lambalgen, M. (2001). Semantics as a foundation for psychology: A case study of Wason's selection task. *Journal of Logic, Language and Information, 10,* 273–317.

Stenning, K., & van Lambalgen, M. (2004). A little logic goes a long way: Basing experiment on semantic theory in the cognitive science of conditional reasoning. *Cognitive Science, 28,* 481–529.

Stenning, K., & van Lambalgen, M. (2005). Semantic interpretation as computation in nonmonotonic logic: The real meaning of the suppression task. *Cognitive Science, 29,* 919–960.

Stenning, K., & van Lambalgen, M. (2008). *Human reasoning and cognitive science.* Cambridge, MA: MIT Press.

Tan, M. T., Tian, G.-L., & Ng, K. W. (2010). *Bayesian missing data problems: EM, data augmentation and noniterative computation.* Boca Raton, FL: Chapman & Hall/CRC Biostatistics Series.

Verguts, T., & Boeck, P. D. (2002). The induction of solution rules in Raven's Progressive Matrices Test. *European Journal of Cognitive Psychology, 14,* 521–547.

Wason, P. C. (1966). Reasoning. In B. M. Foss (Ed.), *New horizons in psychology* (pp. 135–151). Baltimore, MD: Penguin Books.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13,* 103–128.

Zeileis, A., Hornik, K., & Murrell, P. (2009). Escaping RGBland: Selecting colors for statistical graphics. *Computational Statistics and Data Analysis, 53,* 3259–3270.

Received October 27, 2009
Revision received September 13, 2010
Accepted October 25, 2010 ■