THINKING & REASONING, 2008, 14 (3), 281–303

Psychology Press
Taylor & Francis Group

Matching bias in the selection task is not eliminated by explicit negations

Christoph Stahl and Karl Christoph Klauer
University of Freiburg, Germany

Edgar Erdfelder
University of Mannheim, Germany

The processes that guide performance in Wason's selection task (WST) are still under debate. The matching bias effect in the negations paradigm and its elimination by explicit negations are central arguments against a substantial role for inferential processes. Two WST experiments were conducted in the negations paradigm to replicate the basic finding and to compare effects of implicit and explicit negations. Results revealed robust matching bias in implicit negations. In contrast to previous findings, matching bias was reduced but not eliminated in conditions using explicit negations. Model-based analyses suggest that matching bias is due to a switch towards a negative test strategy caused by negations.

Keywords: Reasoning; Heuristics; Wason selection task; Matching bias.

Some 40 years ago, Peter Wason (1966) devised a task that has come to be known as the four-card selection task or as the Wason selection task (WST). In this task reasoners are presented with cards that have a number on one side and a letter on the other side, and a rule is introduced such as, "If there is an A on the letter side, then there is a 3 on the number side". Four cards are shown that represent instances of the antecedent and the consequent as well as instances of their negation on the visible sides. For example, the four

Correspondence should be addressed to Christoph Stahl, University of Freiburg, Institute for Psychology, D-79085 Freiburg, Germany. E-mail: stahl@psychologie.uni-freiburg.de

We are grateful to Valerie Thompson and three anonymous reviewers for their thoughtful comments on an earlier version of the manuscript. The research reported in this paper was supported by grant Kl 614/31-1 from the Deutsche Forschungsgemeinschaft to the second author.

http://www.psypress.com/tar DOI: 10.1080/13546780802116807

cards might show A, B, 3, and 4. Reasoners are to decide which cards would have to be turned in order to test whether the rule is true or false. In the following we refer to cards representing instances of the antecedent and the consequent as p and q, and to cards representing their negations as ~p and ~q, respectively.

The WST is typically solved correctly only by a small minority of participants (for recent reviews, see Evans, Newstead, & Byrne, 1993, ch. 4; Evans & Over, 2004, ch. 5; Oaksford & Chater, 2003). Assuming that the above rule is understood as a conditional in which the antecedent is sufficient for the consequent, the logically correct response is to select the card showing an A (i.e., the p card) and the card showing a 4 (i.e., the ~q card). Only the p and ~q cards can reveal a violation of the rule when turned over. However, they are selected only by a small proportion of participants, typically fewer than 10%, whereas by far the most frequent choices are to select the card with an A (i.e., the p card alone) or the two cards showing A and 3 (i.e., the p and q cards).

The low proportion of normatively correct responses fuelled a debate on whether participants' choices in the WST are rational or not. One possibility is that the participants' interpretation of the task differs from the experimenter's, and that responses are logically correct if the interpretation of the task is taken into account (e.g., Ahn & Graham, 1999; Gebauer & Laming, 1997; Margolis, 1987, ch. 8). Another possibility is that logic is not the appropriate normative yardstick. For example, the WST has been seen as a problem of optimal data selection (ODS) in a Bayesian framework (Oaksford & Chater, 1994; see also Nickerson, 1996), accounting for participants' most frequent selections as normatively correct in that framework. Still another position is that participants often do not reason at all in the WST, but let themselves be guided by preconscious heuristics that endow some of the cards with subjective relevance (Evans, 1995), heuristics that may nevertheless reflect a kind of rationality that ensures the successful attainment of goals in everyday life (Evans & Over, 1996). Theories of the WST (Evans, 1995; Evans & Over, 2004, ch. 9; Johnson-Laird, 1995; Oaksford & Chater, 1994; Sperber, Cara, & Girotto, 1995, among others) have taken different stances on this issue.

MATCHING BIAS IN THE NEGATIONS PARADIGM

Any theory of the WST should be able to account for the dramatic effects on selection behaviour that occur in the so-called negations paradigm. In a classic study Evans and Lynch (1973) presented rules with negated antecedents or consequents. There were four conditions with the following representative examples of rules (note that the original wording in Evans & Lynch, 1973, was slightly different):

- A3: If there is an A on the letter side, then there is a 3 on the number side.
- An3: If there is an A on the letter side, then there is not a 3 on the number side.
- nA3: If there is not an A on the letter side, then there is a 3 on the number side.
- nAn3: If there is not an A on the letter side, then there is not a 3 on the number side.

The major finding, replicated since then several times (Evans et al., 1993, ch. 4), was that across the four rules each card, p, ~p, q, and ~q, was selected more frequently if the number or letter on it was mentioned in the rule; that is, if the card matched the letter or number in the rule. For example, in the An3 condition the proportion of (p, ~q) card selections (i.e., the A and 3 cards) was much higher than the typical 5–10% of correct solutions. Note that we continue to refer to the antecedent and consequent of the shown rule by p and q, respectively, regardless of whether these are affirmative or negative propositions.

In a review of this so-called matching-bias phenomenon, Evans (1998, see also Evans & Handley, 1999) argued that the effect is tied locally to the negations in the rule rather than to the global interpretation of the rule. Negation supposedly leads to greater subjective relevance and thereby to more selections of the matching cards than of the mismatching cards. Importantly, relevance judgements, as well as consequent card selections, are thought to be made for each card separately and independently. More generally, Evans' (1995) initial account by relevance argues that, in the WST, relevance alone determines responses, and that reasoning processes do not contribute to performance in this task.

Support for this interpretation is drawn from a study that reports the elimination of matching bias. Evans, Clibbens, and Rood (1996) used selection tasks with so-called explicit negations in which the negative cards bore generic descriptions of an exemplar of the contrast set. For example, the ~p-card would show "A letter which is not A" on its visible side. One major finding by Evans et al. was that matching bias in the WST was eliminated when explicit negations were used. Evans et al. argued that it is the reduced subjective relevance of cards with implicit negations that causes the matching bias in the WST; for example, participants simply do not realise that a 4 has any logical relation to the rule "If there is an A on the letter side, then there is a 3 on the number side". This difficulty is removed through the introduction of explicit negations, and along with it matching bias disappeared in the studies by Evans et al. (1996).

In a somewhat similar vein, Sperber et al. (1995) have argued that introducing negations decreases the mental effort that has to be invested in thinking about negated cases on the visible and invisible sides of the cards. It should be noted, however, that they only make this argument explicitly for the An3 type of rule. If negations on the cards are explicit, on the other hand, it should take relatively little effort to identify and think of negated as well as affirmative cases on the cards, irrespective of whether there are negations in the rule or not. This suggests that Sperber et al.'s (1995) account by relevance might also be able to account for an absence of matching bias in the WST with explicit negations on the cards.

The elimination of matching bias is theoretically important, because it poses problems for almost any account of the WST and of matching bias in it other than the above accounts by relevance. For example, Margolis (1987, ch. 8) felt that the asymmetrically negated rules (An3 and nA3; e.g., "If there is an A, then there is not a 3") do not as easily invite their converses (e.g., "If there is not a 3, then there is an A") as the symmetrical rules (A3 and nAn3). However, it is difficult to see why effects of negations on rule interpretation and invited inferences should be affected by the nature of the negations (implicit versus explicit) on the cards.

Similarly, Oaksford and Stenning (1992, p. 842) argue that introducing negations alters the interpretation of the rule and the processes of reasoning about the rule so as to create the matching phenomenon for rules with negated antecedents. In their view, reasoners attempt to interpret the rule by constructing the pairings that the rule prescribes. For example, for the rule "If not A, then 3", participants first identify the named elements, then construct the class of elements specified by the negated antecedent (that is, all letters other than A), and then pair these elements with the consequent; that is, with the number 3. In this way all letters other than A are paired with the number 3. Only the letter A has not been paired with 3. Participants are therefore argued to erroneously conclude that the card with the letter A is excluded by the rule from pairing with 3. Cards with an A on the letter side and a 3 on the number side are thereby falsely identified as falsifying the rule, prompting for the rule "If not A, then 3" the selection of both matching cards (~p,q). In this analysis matching is a configural phenomenon, leading to the representation of a believed violating card (in terms of both its number and letter side) when negations are present in the antecedent. Again, this reasoning strategy should not be disrupted through the introduction of explicit negations on the cards.

Oaskford and Chater (1994) maintain that negations alter the perceived rarity of events. In particular, negated antecedents or consequents are perceived as much more frequent events than non-negated antecedents or consequents. According to their account by optimal data selection, perceived rarity is a major factor in determining card selections and in accounting for matching bias. As discussed by Oaksford (2002), because perceived rarity is unlikely to be affected by the nature of the negations on the card, the elimination of matching bias by explicit negations is again difficult to reconcile with this account.

Based on their suppositional account of the meaning of "if", Evans and Over (2004, ch. 9) proposed that one contribution to matching bias is given by what amounts to a switch from a positive test strategy for testing affirmative conditionals to a negative one for testing conditionals with negated components. In a negative test strategy reasoners proceed from the negation of the focal hypothesis and test it in the same manner as they would test the original focal hypothesis in a positive test strategy (Klayman & Ha, 1987). For example, in the positive test strategy participants test predictions derived from the shown rule, and if no violating card is found, the rule is considered true. In a negative test strategy participants test predictions derived from the negation of the shown rule, and if no card violating these predictions is found, the negation is considered true and the shown rule false. According to Evans and Over (2004, ch. 9), the negation of the rule "If X, then Y" is, in the understanding of most reasoners, the rule "If X, then not Y" (Handley, Evans, & Thompson, 2006; Pollard & Evans, 1980). For example, confronted with the rule "If A, then not 3", participants set out to test what they perceive as its negation, namely the rule "If A, then 3". If they do so in the same manner as they would do for the explicitly shown rule in positive testing, the doubly matching combination ($p, \sim q$) should be selected frequently, because it amounts to the combination (p,q) in relation to the negated rule. Thus, according to the suppositional account proposed by Evans and Over (2004, ch. 9), negated components in the rule trigger a switch from a positive to a negative test strategy, causing matching bias. Again, it is difficult to see why this switch should be blocked by explicit negations on the cards. Matching bias is thereby partly caused by a switch in reasoning strategy, and this component of matching bias should not be eliminated by explicit negations on the cards.

To summarise, the elimination of matching bias by explicit negations is a piece of evidence that is difficult to account for by most accounts of WST performance other than Evans' (1995) and perhaps Sperber et al.'s (1995) account by relevance. Although many studies have replicated the basic matching-bias phenomenon, there does not appear to be a single replication of its elimination by explicit negations. Given the relatively small sample size of the original demonstration (Evans et al., 1996) and the prominent theoretical importance of the elimination, we felt that a replication based on a larger sample was highly desirable, and the attempt to replicate was the primary purpose of the present studies.

A second purpose was to see how matching bias would map on the inference-guessing model of the WST that we recently proposed (Klauer, Stahl, & Erdfelder, 2007). The inference-guessing model is a quantitative specification of Evans' (2006) heuristic-analytic model of reasoning applied to the WST that accounts not only for individual card selection frequencies, but for the frequencies of all $2{\cdot}2{\cdot}2{\cdot}2 = 16$ possible patterns of card combinations that can be selected in principle. It does so through a number of psychologically interpretable parameters that specify rule interpretation and inferential processes. The inference-guessing model comprises two submodels, that is, (a) the inference submodel that describes configural card selection patterns based on inferential processes, and (b) the so-called guessing submodel that describes independent card selections based on processes such as relevance assessments or plain guessing.

The inference submodel assumes that participants arrive at different interpretations of the rule, inviting different inferences, and that they differ with regard to the strategies they use to test these invited inferences. It is assumed that each combination of interpretational and inferential processes is connected to a specific configural pattern of card selections (e.g., selecting p and q, but not ~p and ~q).

The guessing submodel includes parameters for the four card types p, ~q, q, and ~q that represent the probabilities with which each card type is selected. Importantly, these card selections are independent, meaning that the probability of selecting, for example, the q card is not affected by whether the p card, or any other card, has been selected or not. A summary of the model's parameters is given in the Appendix, Table A1 (see Klauer et al., 2007, for additional detail on the inference-guessing model).

Klauer et al. (2007) have shown that, across a series of experiments, about 75% of the data were described by the inference submodel, whereas only about 25% of the data could be accounted for by the guessing submodel. This finding suggests that independent card selections can account only for a small proportion of the data, and that card selection is based on processes other than card-wise independent relevance judgements for the majority of participants.

The fact that most card selections in the WST are not independent does not per se exclude the possibility that the matching phenomenon may be based on independent card selections as captured by the guessing submodel. However, typical data from the negations paradigm suggest that the matching phenomenon is not based on local, independent effects on card selections. Consider the *not*-heuristic, which predicts increased subjective relevance of the ~q card under the An3 rule as compared to the A3 rule. This increased relevance should render selection of the ~q card more likely, an effect that is assumed to act locally and independently of the selection or non-selection of other cards. Thus, the *not*-heuristic predicts that negation of the consequent should lead to increased selection frequencies for all patterns that include the ~q card. However, typical data demonstrate an increase in selection frequencies of only the (p, ~q) pattern, whereas other patterns that include the ~q card (for example, the ~q card alone), are not selected more frequently. In some cases, selections are even *less* frequent. In other words, in a typical negation experiment, the increase in selections of the ~q card from rule A3 to rule An3 is not independent but interacts with, among others, the selection of the p card.

It thus appears unlikely that the matching phenomenon is solely based on independent card selections. However, it remains perfectly possible that independent card selections based on relevance heuristics contribute to the matching effect. Model-based analyses can provide further insight into the processes underlying the matching phenomenon. In particular, they provide a means to evaluate whether matching bias is based on independent card selections, as suggested by the relevance account. If matching bias primarily follows perceived relevance of the individual cards, the effects of negations in the negations paradigm should be seen in the parameters of the guessing submodel. In contrast, configural card selections are expected if the matching phenomenon reflects changes in how the different rules with negated components are interpreted or in the inferential processes that they elicit. If this is the case, the effects of negations should be seen in the inference part of the inference-guessing model. Above and beyond this fundamental distinction, model-based analyses can help pinpoint the specific interpretational and inferential processes that, taken together, can account for the matching phenomenon.

In sum, the relevance account of matching bias suggests that matching is based on independent card selections, and predicts that it is eliminated by explicit negations on the cards. In contrast, while the other theories of the WST differ with regard to the independence of card selections, an elimination of matching by explicit negations is inconsistent with most of them.

In the present research two experiments in the negations paradigm were conducted that are reported and discussed together. Experiment 1 used the classical negations paradigm with four groups corresponding to the four rules A3, An3, nA3, and nAn3; negations on the cards were implicit. For example, the ~p card for the rule "If there is an A on the letter side, then there is a 3 on the number side" might be a card with the letter B on it.

In Experiment 2 there were eight conditions, four of them using a negations paradigm with implicit negations and the remaining four with explicit negations. For further reference, the eight resulting groups are labelled IA3, IAn3, InA3, InAn3, EA3, EAn3, EnA3, and EnAn3, where "I" stands for implicit negations and "E" for explicit negations.

METHOD

The studies reported in this paper were implemented as World Wide Web (WWW) experiments. Each participant performed only one WST, and experimental manipulations were implemented between participants. In Experiment 1 four conditions were created, corresponding to the four possible combinations of negations of the antecedent and the consequent of the rule (i.e., conditions A3, An3, nA3, and nAn3). In Experiment 2 eight conditions were created: four conditions implemented a negations paradigm using implicit negations (labelled IA3, IAn3, InA3, and InAn3), and four additional conditions implemented a negations paradigm using explicit negations (labelled EA3, EAn3, EnA3, and EnAn3).

Participants were randomly assigned to the different experimental groups. The experiments were advertised in several newsgroups, submitted to various search machines, and publicised in several WWW documents that collect links to on-line studies and experiments. The experiment was described as a short logic test with individualised feedback, conducted for scientific purposes. The experiment was offered in a German and an English version that were reached by different links.

Participants

Participants were sampled via the Internet. In Experiment 1 mean age was $M = 26.4$ years, 43% of participants were male, and 19% participated in the German version. In Experiment 2 mean age was $M = 28.6$ years, 44% of participants were male, and 9% participated in the German version. In Experiment 1 there were 336, 300, 341, and 349 participants in the groups labelled A3, An3, nA3, and nAn3, respectively. In Experiment 2 there were 360, 332, 329, 321, 346, 300, 205, and 323 participants in the groups labelled IA3, IAn3, InA3, InAn3, EA3, EAn3, EnA3, and EnAn3, respectively.

Instructions and procedure

The experiment consisted of a start page, an experimental page, and a feedback page. The start page asked participants whether they would like to participate in a short scientific study about reasoning of a duration of about 5 minutes. They were also asked to read the instructions carefully if they were willing to participate. Persons wishing to proceed to the problem indicated their intent by clicking on a link labelled “yes” on the start page. For each such participant an experimental page was generated online. The program generating the experimental page randomly assigned the participant to one of the experimental groups and selected random letters from the alphabet (excluding the letters I, O, and V because of their similarity to numerals) and integer numbers between 1 and 9 to be used in the Wason selection problem. The experimental page comprised the instructions, the Wason selection problem, and a biographical questionnaire. In the questionnaire participants were asked for demographic bits of information about themselves. Additional questions addressed the participant's language proficiency, prior experience with the just-completed task or similar card selection tasks, and whether the participant "had answered all questions carefully and participated for the first time" or whether he or she "just want[s] to see the results by way of trial without seriously participating in the study". These questions were used to screen out potentially suspicious data sets as explained next. The feedback page provided feedback about the participant's selection and the normatively correct selection. The rationale for the normative selection (p, not-q) was explained.

A number of studies have discussed potential problems and advantages of Internet research (e.g., Kraut et al., 2004; Reips, 2002). In the present context, important problems are the low experimental control over the participants' situational circumstances and behaviour, the problem of possible multiple participation, and the potential problem of selective dropout. Selective dropout is a problem if dropout affects some of the experimental groups more strongly than others, thereby compromising the comparability of the experimental groups. Several techniques have been proposed to minimise such problems (Reips, 2002). Following these recommendations, submissions were accepted for data analysis in the present studies only if (a) no submission had been previously received from the same Internet protocol (IP) address. For this purpose, a cumulative record was kept of the IP addresses of submissions throughout the present experiments to screen out any participant who might already have participated in the current or a previous experiment in the series. Furthermore, participants' data were accepted only if they stated (b) that they had responded to all questions carefully and submitted data for the first time, and (c) that they were not familiar with the problem or similar card selection problems. Finally, (d) participants were excluded if they stated that their English (or German in the German version) was poor. These measures aimed at minimising the potential problems of multiple participation, lack of seriousness, motivation, and comprehension. A given experiment remained online until there were at least 300 participants who fulfilled the above criteria in each experimental group.

In Experiment 1 the experimental page began with the following standard instruction: "Below you see a number of cards from a set of cards. Each card in the set has a capital letter on one side and a digit on the other. Naturally, only one side is visible in each case. For the set of cards, a rule has been stated. It is: ..." This was followed by a rule with a randomly sampled capital letter in the antecedent and a randomly sampled number between 1 and 9 in the consequent, and negations of antecedent and consequent added depending on condition. For example, in the An3 condition, a possible rule would be: "If there is an A on the letter side of the card, then there is not a 3 on the number side". Another possible rule in the same condition would be: "If there is a K on the letter side of the card, then there is not a 9 on the number side".

In the next paragraph participants were informed that "you must decide which card(s) displayed would have to be turned over in order to test the truth or falsity of the rule. Please use the mouse to check the card(s) that would have to be turned over. Do not check cards that would not have to be turned. You may take as long as you like."

Below this, four cards were displayed in a row. Letter sides showed a capital letter in black on a white card; number sides a number in black on a grey card. The four cards displayed the letter mentioned in the rule, another randomly sampled letter, the number mentioned in the rule, and another randomly sampled number in random order. Below each card a box could be checked to signal selection of the card. No action was required if a card was not to be selected. All randomisations were carried out anew for each participant.

Evans et al. (1996) carefully discussed how to introduce explicit negations in the WST in their Experiment 3, and we followed their procedures relatively closely for Experiment 2. In particular, in Experiment 2 participants were instructed that the cards described letter–number pairs: "Each card represents a letter–number pair. Each card has information about the letter from the pair on one side and information about the number on the other side. Naturally, only one side is visible in each case. For the letter–number pairs, a rule has been stated. It is: ..."

In Experiment 2 the four cards had descriptions written on them instead of just certain letters or numbers. In the implicit groups these were always in the form "the letter B", "the number 2", and so forth; in the explicit groups negated cases were represented on the cards as "a letter which is not C" or "a number which is not 4", and so forth.

RESULTS

To assess the possibility of selective dropout, we tested as a first step whether the numbers of accepted submissions were significantly different between the experimental groups. The number of participants did not differ significantly between the different experimental groups in each experiment, $\chi^2(3) = 4.25$, $p = .24$, and $\chi^2(7) = 7.86$, $p = .35$ for Experiment 1 and 2, respectively. We therefore proceeded to quantify the matching bias effect and compare its magnitude in the implicit and explicit groups, using the indices suggested by Evans et al. (1996). Finally, we used model-based analyses to investigate to which extent the matching phenomenon can be explained by independent card selections. The model-based analyses also allowed us to identify processes of interpretation and inferential reasoning that, taken together, can account for the matching phenomenon.

Matching and logic indices

Table 1 gives the selection frequencies for individual cards. For a traditional analysis of these frequencies, following Evans et al. (1996), we computed an antecedent matching index (AMI), a consequent matching index (CMI), and a logic index (LI; cf. Pollard & Evans, 1987). The AMI is calculated by counting the number of selections of matching letter cards minus the number of selections of mismatching letter cards. The CMI is calculated analogously on the basis of the number cards, and the LI is computed by adding 1 for a correct selection (i.e., a selection of p or ~q) and subtracting 1 for each selection of the remaining cards. Because the numbers of participants differed slightly between groups, an average index was computed for each group separately, and these averages were then averaged across groups. This ensured that each group had equal weight in the grand mean.[1] The results are given in Table 1.

In Experiment 1, replicating previous work (Evans et al., 1996), both AMI and CMI were significantly larger than zero ($t = 7.57$ and 12.33, respectively, $df = 1325$, both ps $< .001$). Thus, the usual matching effect

TABLE 1
Frequencies of individual card selections, as well as matching and logic indices, for Experiments 1 and 2

Exp.	*Negation*	*Group*	*p*	*~p*	*q*	*~q*	*AMI*	*CMI*	*LI*
1	Implicit	A3	**243**	57	**169**	63	.19	.25	.43
		An3	**231**	40	53	**168**			
		nA3	215	**111**	**155**	113			
		nAn3	181	**140**	116	**175**			
2	Implicit	IA3	**241**	65	**177**	91	.18	.28	.40
		IAn3	**255**	62	63	**152**			
		InA3	207	**119**	**157**	108			
		InAn3	154	**108**	105	**149**			
	Explicit	EA3	**225**	93	**127**	89	.12	.17	.32
		EAn3	**229**	91	106	**137**			
		EnA3	197	**79**	**137**	88			
		EnAn3	197	**128**	137	**143**			

Matching card selections are set in boldface. Exp. = Experiment, AMI = Antecedent Matching Index, CMI = Consequent Matching Index, LI = Logic Index (see text).

[1]This was not ensured in the statistical analysis via *t*-tests, which were chosen for reasons of comparability with Evans et al. (1996), but we also ran an analysis based on the frequencies aggregated across participants per group in which the different groups were weighted equally. This analysis yielded the same pattern of significant and non-significant results as reported in the text.

appeared on both antecedent and consequent cards. Unlike in the Evans et al. (1996) studies, the LI was also relatively large and significantly larger than zero ($t = 14.25$, $p < .001$).

In the implicit groups in Experiment 2, as expected, the AMI, CMI, and LI were significantly larger than zero ($t = 8.11$, 14.62, and 13.54, respectively, $df = 1341$, all ps $< .001$). Importantly, and in contrast to the findings reported by Evans et al. (1996), the indices were also significantly larger than zero in the explicit groups ($t = 5.14$, 7.90, and 10.30, respectively, $df = 1273$, all ps $< .001$). The differences between implicit and explicit groups in the AMI, the CMI, and the LI were significant: for the difference in the AMI, $t = 2.08$, $p = .04$; in the CMI, $t = 4.00$, $p < .01$; in the LI, $t = 2.09$, $p = .04$ ($df = 2614$).

Model-based analyses

Going beyond the individual card selection frequencies given in Table 1, the model-based analyses reported below are based on the frequencies of all 16 possible card selection patterns (see Table 2). Using the HMMTree software (Stahl & Klauer, 2007), we fitted the inference-guessing model (Klauer et al., 2007) separately to the three full negations paradigms: (1) the data from Experiment 1, (2) the data from the implicit groups of Experiment 2, and (3) the data from the explicit groups of Experiment 2.[2] In a first set of analyses we tested whether the guessing submodel alone was capable of fitting the data. This was not the case, $G^2 = 691.33$, $df = 44$, $G^2 = 729.82$, $df = 44$, and $G^2 = 787.98$, $df = 10$, for the three data sets, respectively, all $p < .001$, and this finding further supports the conclusion that the majority of card selections in the WST are not based on card-wise independent processes (Klauer et al., 2007). Next we fitted the full inference-guessing model. Compared to the guessing submodel alone, model fit was significantly improved, but it was still not satisfactory ($G^2 = 23.46$, $df = 12$, $p = .02$; $G^2 = 33.39$, $df = 12$, $p < .01$; $G^2 = 61.01$, $df = 12$, $p < .01$). However, when the inference submodel was extended by an additional parameter that models the probability of a switch to a negative test strategy, as suggested by Evans and Over (2004, ch. 9), model fit was good, $G^2 = 6.42$, $df = 9$, $p = .70$, $G^2 = 15.50$, $df = 10$, $p = .11$, and $G^2 = 15.16$, $df = 10$, $p = .13$, for the three data sets, respectively. The extended inference-guessing model thus provided a full quantitative account of the present data. Parameter estimates are given in the Appendix, Table A2.

[2]Additional detail regarding the model-based analyses, including the input files required to fit the inference-guessing model to the observed frequencies of Experiments 1 and 2 using the HMMTree program, can be requested from the first author.

TABLE 2
Frequencies of card selection patterns for Experiments 1 and 2

Exp.	*Group*	*0000*	*0001*	*0010*	*0011*	*0100*	*0101*	*0110*	*0111*	*1000*	*1001*	*1010*	*1011*	*1100*	*1101*	*1110*	*1111*
1	A3	10	10	35	5	11	16	4	2	99	12	**105**	3	8	1	1	14
	An3	8	21	12	2	10	2	12	2	76	**124**	10	7	4	2	0	8
	nA3	12	19	20	7	16	7	**42**	3	79	42	37	14	10	1	12	20
	nAn3	8	40	11	7	26	**64**	11	1	54	29	53	7	9	3	2	24
2	IA3	6	18	43	6	7	22	6	2	92	20	**99**	2	7	2	0	19
	IAn3	9	20	12	3	15	9	20	0	110	**111**	15	1	6	0	4	8
	InA3	9	11	21	6	17	19	**43**	6	74	40	49	10	12	0	6	16
	InAn3	9	34	17	7	22	**56**	8	1	39	34	55	5	8	1	1	11
	EA3	14	8	10	1	9	43	9	7	88	22	**90**	0	15	0	2	8
	EAn3	12	18	19	3	17	24	25	1	84	**76**	42	3	10	1	2	11
	EnA3	5	18	20	8	17	13	**20**	2	62	36	71	1	12	0	5	10
	EnAn3	8	25	22	2	19	**56**	16	1	53	33	70	5	13	2	2	19

Selection patterns are represented by combinations of the symbols 0 and 1, representing selection of the p-card, the ∼p-card, the q-card, and the ∼q-card, respectively, where 1 indicates selection and 0 indicates nonselection (e.g., 1001 represents the selection of the p and ∼q cards). Matching patterns are set in boldface. Exp. = Experiment. I = Implicit, E = Explicit.

Although the guessing submodel alone was not able to account for the data, the parameter estimates of the extended inference-guessing model revealed that, consistent with previous findings (Klauer et al., 2007), about 25% of responses were governed by independent card selections. We were interested in whether there was evidence for matching bias in these responses. Thus, we analysed whether the parameters modelling independent card selections were involved in accounting for the matching phenomenon (i.e., whether they were affected by rule type in a manner consistent with matching). Overall, there was little evidence for effects of rule type on the parameters of the guessing submodel (i.e., only one out of 12 tests for rule type effects on the guessing parameters was significant at $\alpha = .05$; cf. Table A2). Thus, it appears that independent card selections did not contribute substantially to the matching phenomenon.

In contrast, there were strong and systematic effects of rule type on the parameters of the inference submodel that governs configural card selections. In all three data sets asymmetrically negated rules (An3 and nA3) elicited a large proportion of switches to a negative test strategy (parameter *n;* cf. Table A2). Also in all three data sets the perceived direction of the rule (i.e., whether the rule invites inferences from letters to numbers or from numbers to letters; parameter *d*) was affected by rule type, although effects were significant only in the two data sets from Experiment 2. Moreover, the interpretation of p as necessary versus sufficient for q (parameters *sl and sln*) was also affected by rule type across all three data sets. We conclude that introducing negations into the rule can be said to affect processes of rule interpretation and inferential reasoning that govern configural card selections. In other words, the extended inference-guessing model accounts for the matching phenomenon by changes across rule type groups in the interpretation of the rule as well as in inferential processing.

DISCUSSION

Replicating previous findings, matching indices revealed robust matching bias in the two data sets implementing implicit negations (Experiment 1 and the implicit groups from Experiment 2). In contrast to previous findings, however, explicit negations reduced but did not eliminate matching in Experiment 2. Also in contrast to Evans et al.'s (1996) results, logic indices were larger than zero across all conditions.

Note that the level of matching observed in the implicit conditions of the present data is closely comparable to that observed in Evans et al. (1996). Participants in the present experiments performed only a single WST, and AMI and CMI therefore ranged between -1 and $+1$, whereas participants in Evans et al. worked on four WST tasks, with AMI and CMI ranging from -4 to $+4$. When the present indices are rescaled to range between -4

and +4, an AMI value of .76 results both for Experiment 1 and for the implicit groups of Experiment 2, which is comparable to the value of .73 that was reported by Evans et al. (1996, Exp.3); similarly, the rescaled CMI values were 1.00 and 1.13, respectively, for Experiment 1 and the implicit groups from Experiment 2, which is comparable in magnitude to the value of 1.33 reported by Evans et al. (1996, Exp. 3).

What is the cause of the discrepancy in matching in the conditions with explicit negations in the present data and in Evans et al.'s (1996) data? A simple explanation is that there may have been some residual matching in the explicit negation condition reported by Evans et al. (1996, Exp.3), but that the sample size may not have been large enough for the effect to become statistically significant. We assessed this possibility with a power analysis (performed with G*Power 3; Faul, Erdfelder, Lang, & Buchner, 2007). In the explicit negation conditions of the present Experiment 2, the AMI effect was small in terms of Cohen's (1998) effect-size conventions with an effect size of $d = .14$. To detect an effect of this magnitude with $\alpha = .05$ and a power of at least $1 - \beta = .80$, a sample size of $N = 312$ is required. Given the $N = 32$ reported by Evans et al. (1995) for the explicit conditions, the power to detect an effect of $\mathrm{d} = .14$ was only $1 - \beta = .19$. Similarly, the CMI effect was small, $\mathrm{d} = .22$, in the explicit negation conditions of the present Experiment 2; to obtain such an effect with $\alpha = .05$ and a power of at least $1 - \beta = .80$, a sample size of $N = 129$ is required. Given the $N = 32$ reported by Evans et al. (1995) for the explicit conditions, the power to detect an effect of $\mathrm{d} = .22$ was only $1 - \beta = .34$. It is seen that the residual matching effects in the explicit conditions of the present Experiment 2 were of small magnitude ($\mathrm{d} \leq .22$; cf. Cohen, 1988), and that chances were small to detect them given a sample of $N = 32$, in which only effects of medium size ($\mathrm{d} \geq .45$) or larger are detected with an acceptable power of $1 - \beta = .80$.

The second discrepancy between the present data and Evans et al.'s (1996) results concerns the logic index. It was significantly larger than zero in our data, but not significant in Evans et al.'s data. This may again reflect the difference in test power between the studies. Another possibility is suggested by the fact that WST performance is known to depend on the level of cognitive ability (Newstead, Handley, Harley, Wright, & Farrelly, 2004; Stanovich & West, 1998). Participants in Evans et al.'s (1996) Experiment 3 were described as students attending psychology courses as part of a nursing degree programme at the University of Plymouth, whereas the educational and occupational status of participants in the present study may have been higher; for example, participants reported having spent an average of 13 years at school (including college/university). Thus, the differences in the logic index may reflect differences in the level of cognitive ability between the present and the previous participant samples.

Summing up, the traditional analysis revealed a robust matching effect even in the groups with explicit negations. Explicit negations reduced the matching indices significantly to approximately two-thirds of the values obtained in the groups with implicit negations, but the remaining evidence for matching bias was still strong. As argued by Evans et al. (1996) and Evans (1995), whether matching bias is eliminated by explicit negations or not has important theoretical implications. As elaborated in the Introduction, an elimination of matching by explicit negations is difficult to reconcile with most accounts of the WST and of matching bias in particular (Evans & Over, 2004; Oaksford, 2002; Oaksford & Chater, 1994; Oaksford & Stenning, 1992; Margolis, 1987). The present findings suggest that the difficulty may not exist in the first place. Instead, the present data provide tentative support to some of these accounts, as we will point out after we have discussed the model-based analyses.

The fact that a robust matching effect remained evident in the explicit groups of Experiment 2 is difficult to reconcile with the relevance account. The present data thereby do not support an account of the matching phenomenon that relies solely on local relevance judgements affected by a *not*-heuristic as suggested by Evans (1995). To illustrate, consider the first two rows in Table 2, which refer to the rules A3 and An3, and focus on the columns labelled 0001 (representing selection frequencies of the ~q card alone) and 1001 (representing selection frequencies of the p card together with the ~q card). The *not*-heuristic predicts increased subjective relevance of the ~q card under the An3 rule as compared to the A3 rule, which should render selection of that card more likely, independent of the selection or non-selection of other cards. In other words, the *not*-heuristic predicts that selection frequencies should increase from the A3 to the An3 rule to a similar extent in all of the columns that represent card selection patterns that include the ~q card. Contrasting this prediction, it can be seen in Table 2 that an approximately tenfold increase in ~q card selections is evident for the case in which the p card is also selected (i.e., column 1001), but that this increase is much weaker in case the p card is not selected (i.e., column 0001). In other cases, selection frequencies were even found to *decrease* instead (see the 0011, 0101, and 1111 columns).

Above and beyond their role in the matching effect, the present findings provide further support for the position that an account in terms of relevance heuristics cannot fully explain performance in the WST. While it is perfectly possible that these heuristics do contribute to the matching phenomenon, the present data suggest that their role is limited, and that processes other than such local relevance judgements must underlie the matching phenomenon.

Our central finding—that WST performance, as well as the magnitude of the matching effect, was only slightly affected by explicit negations—is consistent with the idea that there is a central role for reasoning in the selection task. The contribution of reasoning processes is further supported by logic indices that were significantly larger than zero in all three groups. We suggest that the matching phenomenon is, at least in part, a result of the combined operation of interpretive and inferential processes. In the following we will evaluate the ability of two recent quantitative models of the WST to fit the present data and we will show how the inference-guessing model can account for the matching phenomenon in terms of interpretive and inferential reasoning processes.

Model-based analyses

We conducted two sets of model-based analyses, the first based on the ODS model by Oaksford and Chater (2003) that can be seen as a quantitative specification of their account by ODS, the second based on the inference-guessing model sketched above (Klauer et al., 2007) that can be seen as a quantitative specification of Evans' (2006) revised heuristic-analytic model. Both models were fit to three data sets: The response-pattern frequencies of Experiment 1, of the implicit groups in Experiment 2, and of the explicit groups in Experiment 2 (see Table 2 for the frequencies and Klauer et al., 2007, for details on the models). Neither model approached empirically adequate descriptions of the response-pattern frequencies; that is, model fit was poor.

Failure to obtain model fit for a quantitative specification of a theory is only relatively weak evidence against the parent theory, because any quantitative specification relies on auxiliary assumptions that may be responsible for lack of fit. Alternative specifications of the theory, employing other auxiliary assumptions, may still exist that provide better fits. For example, in the case of the ODS model by Oaksford and Chater (2003), one auxiliary assumption is that card selections are independent from each other, and it is possible to relax that assumption to accommodate dependencies in card selections. Klauer et al. (2007) considered versions of the ODS model that relaxed the independence assumption, but found that the modified models still failed to describe the response-pattern frequencies adequately. Nevertheless, it is perfectly possible that alternative modifications might be found that provide better fits of the data.

In the case of the inference-guessing model we modified the model by integrating Evans and Over's (2004, ch. 9) idea that negations in the rule trigger a switch from a positive to a negative test strategy. The idea is that negations in the rule cause reasoners to attempt to test what they perceive to be the negation of the rule. Adding this principle to our model allowed us to account for the pronounced changes that occur from rule to rule in the negations paradigm adequately. More specifically, we added a new parameter to the inference-guessing model to describe the likelihood with which participants switch to a negative test strategy when presented with a rule with negated constituents, as suggested by Evans and Over (2004). As a result of this simple modification, the modified model now yielded adequate descriptions of the three data sets, with model fit indices that no longer indicated significant model violations.

The good model fit implies that the modified inference-guessing model is consistent with the response frequency data for all 16 response patterns, across all three data sets. That is, it is not only capable of explaining, for each of the three data sets, the selection frequencies for the four cards and the response frequencies of a few modal response patterns. Rather, it also accounts for the frequencies of the less frequent patterns and for correlations and higher-order dependencies of any kind between the card selections of all four cards.

Moreover, the effects of rule type were mapped on the model parameters in a parsimonious manner. According to the model-based analyses, the matching phenomenon goes back to the joint action of two principles: (1) a high percentage of negative tests in the asymmetrically negated rules (e.g., "If there is an A, then there is not a 3")[3] and (2) a tendency of rules with negated antecedents (e.g., "If there is not an A, then there is not a 3") to elicit their inverse ("If there is an A, then there is a 3") and/or contrapositive ("If there is a 3, then there is an A"),[4] perhaps through a mechanism involving the construction of contrast classes for negated antecedents as suggested by Oaksford (2002; Oaksford & Stenning, 1992). In addition, there was only weak evidence for matching in the parameters of the guessing submodel that describes independent card selections. With regard to the initial account by relevance (Evans, 1995) we conclude that, although there might be a role for relevance judgements in accounting for matching, this role appears to be restricted to a limited proportion of the data.

What were the reasons for the differences in matching between the implicit and explicit groups in Experiment 2? Model-based analyses comparing the data from the implicit and explicit groups suggest that similar processes of rule interpretation and reasoning governed performance in both groups, and that the differences are quantitative rather than qualitative. Quantitative differences were observed in the proportion of negative tests (i.e., parameter n), which was substantially higher in the implicit groups than in the explicit groups. In addition, a smaller proportion of responses was guided by interpretive and reasoning processes for rules with negated antecedents (nA3, nAn3) in the implicit groups as compared to the explicit groups. This is reminiscent of an effect of negated antecedents on

[3]That is, parameter n was particularly large for such rules.

[4]That is, parameters *sl* and *sln* were particularly small for such rules; see Klauer et al. (2007).

rule comprehension: Ormerod, Manktelow, and Jones (1993) found that rules with negated antecedents were understood more slowly than rules with affirmative antecedents in the negations paradigm. But the confusion that may be reflected in these effects was reduced when negations were explicit.

Failure to obtain model fit for a quantitative specification of a parent theory does not invalidate the parent theory. On the other hand, the positive results obtained for the modified inference-guessing model implies that the parent theory—Evans' (2006) revised heuristic-analytic theory combined with Evans and Over's (2004) suppositional account of matching bias—is consistent with the present data. More specifically, these results imply that matching bias, rather than being based solely on independent card-wise relevance judgements, can be understood as a configural phenomenon resulting from the operation of one or more processes of rule interpretation and inference. The inference submodel of the inference-guessing model (Klauer et al., 2007), extended by a parameter representing the possibility of switching to a negative-testing strategy as suggested by Evans and Over (2004, ch. 9), presents a candidate set of such processes. In addition to identifying a set of processes possibly involved in creating the matching phenomenon, the inference-guessing model also delivers quantitative estimates for these processes, as well as a formal specification of the ways these processes interact to produce the set of all possible card selection patterns.

Two additional results were obtained that are relevant for some of the other accounts of the WST mentioned in the discussion. First, as already mentioned above, rules with negated antecedents ("if there is not an A, then there is not a 3") showed a tendency to elicit their inverse ("if there is an A, then there is a 3"),[5] which is consistent with the idea by Oaksford and Stenning (1992) that the interpretation of a negation involves identifying and activating the set of elements that are implied by the negation of the antecedent. The construction of contrast classes may be the underlying mechanism for this effect on the interpretation of the rule. Additional research that targets this possibility is desirable.

Second, model-based analyses of the present data lend support to Margolis's (1987, ch. 8) hypothesis that asymmetrically negated rules (i.e., An3 and nA3) are less likely to invite their converses as are symmetrically negated rules. As suggested by Margolis, negations affected the probability with which participants interpreted the rule (e.g., "if there is not an A, then there is a 3") as implying its converse ("if there is a 3, then there is not an A").[6] This effect was present in all three data sets, but it was restricted to rules with negated antecedents.

[5]That is, parameter *sl* was particularly small for such rules; see Klauer et al. (2007).

[6]That is, parameter *d* was particularly small for symmetrically negated rules; see Klauer et al. (2007).

CONCLUSIONS

The present research considerably enhances the set of available data relevant for the matching phenomenon in the selection task. It was found that matching is reduced but not generally eliminated by explicit negations. Additional research is required to pinpoint the factors that have caused the present results to diverge from those obtained by Evans et al. (1996), although we argued that the smaller test power inherent in Evans et al.'s (1996) studies may be responsible for the divergent results. In any event, it can no longer be taken as given that explicit negations eliminate matching bias. This is a theoretically important result given that the elimination of matching by explicit negations poses a problem for many accounts of the WST and of the matching phenomenon in it.

Furthermore, model-based analyses revealed that matching bias could not be explained by independent cardwise relevance judgements. Instead, these analyses suggest that the matching phenomenon is the result of a switch to a negative test strategy (Evans & Over, 2004, ch.9), in combination with well-circumscribed effects of negations on interpretational processes.

Manuscript received 13 June 2007
Revised manuscript received 8 January 2008
First published online 19 May 2008

REFERENCES

Ahn, W., & Graham, L. M. (1999). The impact of necessity and sufficiency information in the Wason four-card selection task. *Psychological Science, 10*, 237–242.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Evans, J. St. B. T. (1993). The mental model theory of conditional reasoning: Critical appraisal and revision. *Cognition, 48*, 1–20.

Evans, J. St. B. T. (1995). Relevance and reasoning. In S. E. Newstead & J. St. B.T. Evans (Eds.), *Perspectives on thinking and reasoning* (pp. 147–172). Hove, UK: Lawrence Erlbaum Associates Ltd.

Evans, J. St. B. T. (1998). Matching bias in conditional reasoning: Do we understand it after 25 years? *Thinking and Reasoning, 4*, 45–82.

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

Evans, J. St. B. T., Clibbens, J., & Rood, B. (1996). The role of implicit and explicit negation in conditional reasoning bias. *Journal of Memory and Language, 35*, 392–409.

Evans, J. St. B. T., & Handley, S. J. (1999). The role of negation in conditional inference. *Quarterly Journal of Experimental Psychology, 52*, 739–769.

Evans, J. St. B. T., & Lynch, J. S. (1973). Matching bias in the selection task. *British Journal of Psychology, 64*, 391–397.

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

Evans, J. St. B. T., & Over, D. E. (1996). *Rationality and reasoning*. Hove, UK: Psychology Press.

Evans, J. St. B. T., & Over, D. E. (2004). *If*. Oxford, UK: Oxford University Press.

Faul, F., Erdfelder, E., Lang, A-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191.

Gebauer, G., & Laming, D. (1997). Rational choices in Wason's selection task. *Psychological Research, 60*, 284–293.

Handley, S. J., Evans, J. S. B. T., & Thompson, V. A. (2006). The negated conditional: A litmus test for the suppositional conditional? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 559–569.

Johnson-Laird, P. N. (1995). Inferences and mental models. In S. E. Newstead & J. St. B. T. Evans (Eds.), *Perspectives on thinking and reasoning* (pp. 115–146). Hove, UK: Psychology Press.

Klauer, K. C., Stahl, C., & Erdfelder, E. (2007). The abstract selection task: New data and an almost comprehensive model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 680–703.

Klayman, J., & Ha, Y-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94*, 211–228.

Kraut, R., Olson, J., Banaji, M., Bruckman, A., Cohen, J., & Couper, M. (2004). Psychological research online: Report of board of scientific affairs' advisory group on the conduct of research on the Internet. *American Psychologist, 59*, 105–117.

Margolis, L. (1987). *Patterns, thinking and cognition: A theory of judgement*. Chicago, IL: University of Chicago Press.

Newstead, S. E., Handley, S. J., Harley, C., Wright, H., & Farrelly, D. (2004). Individual differences in deductive reasoning. *Quarterly Journal of Experimental Psychology, 57*, 33–60.

Nickerson, R. S. (1996). Hempel's paradox and Wason's selection task: Logical and psychological puzzles of confirmation. *Thinking and Reasoning, 2*, 1–31.

Oaksford, M. (2002). Contrast classes and matching bias as explanations of the effects of negation on conditional reasoning. *Thinking and Reasoning, 8*, 135–151.

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

Oaksford, M., & Chater, N. (2003). Optimal data selection: Revision, review, and reevaluation. *Psychonomic Bulletin & Review, 10*, 289–318.

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

Oberauer, K. (2006). Reasoning with conditionals: A test of formal models of four theories. *Cognitive Psychology, 53*, 238–283.

Ormerod, T. C., Manktelow, K. I., & Jones, G. V. (1993). Reasoning with three types of conditional: Biases and mental models. *Quarterly Journal of Experimental Psychology, 46*, 653–677.

Pollard, P., & Evans, J. St. B. T. (1980). The influence of logic on conditional reasoning performance. *Quarterly Journal of Experimental Psychology, 32*, 605–624.

Pollard, P., & Evans, J. St. B. T. (1987). Content and context effects in reasoning. *American Journal of Psychology, 100*, 41–60.

Reips, U-D. (2002). Standards for Internet-based experimenting. *Experimental Psychology, 49*, 243–256.

Sperber, D., Cara, F., & Girotto, V. (1995). Relevance theory explains the selection task. *Cognition, 57*, 31–95.

Stahl, C., & Klauer, K.C. (2007). HMMTree: A program for latent-class processing tree models. *Behavior Research Methods, 39*, 267–273.

Stanovich, K. E., & West, R. F. (1998). Cognitive ability and variation in selection task performance. *Thinking and Reasoning, 4*, 193–230.

Wason, P. C. (1966). Reasoning. In B. M. Foss (Ed.), *New horizons in psychology I* (pp. 135–151). Harmondsworth, UK: Penguin.
Wason, P. C., & Evans, J. St. B. T. (1975). Dual processes in reasoning? *Cognition, 3*, 141–154.

APPENDIX

TABLE A1
Parameters of the extended inference-guessing model (see Klauer et al., 2007)

Par.	*Meaning*
a	Probability of response being based on the inference submodel rather than the guessing submodel
Inference Submodel	
c	Probability of a conditional rather than a biconditional interpretation of "if p, then q"
x	Probability of a bidirectional biconditional interpretation (e.g., the rule is interpreted as "if p, then q *and* if q, then p") rather than a case-distinctive biconditional interpretation (e.g., the rule is interpreted as "if p, then q *and* if not p, then not q")
d	Probability of inferences in the forward direction (from letters to numbers) rather than backward direction (from numbers to letters)
sl	Probability of perceived sufficiency ("always if") rather than necessity ("only if") of p for q, given a forward inference
sn	Probability of perceived sufficiency ("always if") rather than necessity ("only if") of q for p, given a backward inference
sln	Probability of perceived sufficiency ("always if") rather than necessity ("only if") of p for q and q for p, given a bidirectional interpretation
i	Probability of irreversible reasoning (inferences only from the visible sides of the cards) rather than reversible reasoning (inferences from visible and invisible sides)
n	Probability of choosing a negative test strategy
Guessing Submodel	
p	Probability of selecting card p in independent card selections
$\bar{p}$	Probability of selecting card ∼p in independent card selections
q	Probability of selecting card q in independent card selections
$\bar{q}$	Probability of selecting card ∼q in independent card selections

Par. = parameter.

TABLE A2
Maximum-likelihood parameter estimates for the extended inference-guessing model, separately for the conditions of Experiments 1 and 2

	Experiment 1				*Experiment 2*							
	Implicit negations				*Implicit negations*				*Explicit negations*			
Par.	*A3*	*An3*	*nA3*	*nAn3*	*A3*	*An3*	*nA3*	*nAn3*	*A3*	*An3*	*nA3*	*nAn3*
p	**.40**	**.59**	.59	.60	**.38**	**.35**	.51	.61	**.21**	**.47**	.33	.70
$\bar{p}$	.30	.23	**.30**	**.25**	.35	.47	**.31**	**.16**	.48	.33	**.49**	**.29**
q	**.40**	.36	**.63**	.39	***.32***	*.40*	***.69***	*.26*	**.38**	.31	**.69**	.30
$\bar{q}$	.42	**.50**	.31	**.48**	.60	**.30**	.32	**.55**	.39	**.38**	.23	**.34**
a	.76	.77	.77	.72	.83	.79	.65	.69	.76	.77	.77	.72
c	.51	.43	.54	.46	.53	.48	.48	.39	.37	.43	.43	.37
x	.94	.99	.95	.93	.93	.98	.92	.93	.92	.96	.89	.97
d	.76	.77	.83	.63	*.66*	*.83*	*.88*	*.56*	*.97*	*.77*	*.73*	*.55*
sl	*.93*	*.89*	*.84*	*.63*	*.96*	*.94*	*.82*	*.54*	*1.00*	*.86*	*.83*	*.66*
sn	.90	.62	1.00	.09	.82	.80	1.00	.35	.72	.71	.83	.53
sln	*.88*	*.91*	*.51*	*.42*	*.85*	*.87*	*.52*	*.49*	*.72*	*.71*	*.83*	*.53*
i	.91	.95	.84	.85	.87	.95	.89	.92	.95	.94	.93	.89
n	*.00*	*1.00*	*.99*	*.08*	*.00*	*.93*	*.66*	*.12*	*.14*	*.63*	*.34*	*.20*

For an explanation of the parameters, see Table A1.
Par. = parameter. Par. values for matching conditions are set in boldface. Par. estimates that differ significantly at $\alpha = .05$ across the four rule types A3, An3, nA3, and nAn3 per condition are set in italics.