THINKING & REASONING, 2006, 12 (2), 214–234

Psychology Press
Taylor & Francis Group

How falsity dispels fallacies

Mary R. Newsome and P. N. Johnson-Laird
Princeton University, Princeton, NJ, USA

From certain sorts of premise, individuals reliably infer invalid conclusions. Two experiments investigated a possible cause for these illusory inference: Reasoners fail to think about what is false. In Experiment 1, 24 undergraduates drew illusory and control inferences from premises based on exclusive disjunctions ("or else"). In one block, participants were instructed to falsify the premises of each illusory and control inference before making the inference. In the other block, participants did not receive these instructions. There were more correct answers for illusory disjunctions whose premises had been falsified than there were for illusory disjunctions that had not been falsified. A second experiment introduced illusory inferences in a real world context that accentuated falsification of premises. Accuracy also improved. Knowledge of how to falsify premises and to consider their implications for true premises transferred to a new problem introduced at the end of the experiment without the falsification instruction. The participants' ratings of the difficulty of the inferences showed that they did not err simply because illusory inferences are perceived to be more difficult than control problems. The model theory predicts these results because it postulates that the limitations of working memory preclude the representation of false information.

"No testimony is sufficient to establish a miracle, unless the testimony be of such a kind that its falsehood would be more miraculous than the fact which it endeavors to establish."
David Hume–Of Miracles, from An Enquiry Concerning Human Understanding.

Correspondence should be address to Mary R. Newsome, Cognitive Neuroscience Laboratory, Baylor College of Medicine, 1709 Dryden Road, Ste. 725, Houston, TX 77030, USA. Email: mnewsome@bcm.tmc.edu

This research was supported in part by a grant to the second author from the National Science Foundation (BCS 0076287) to investigate strategies in reasoning. We thank the members of our laboratory, who have provided much helpful advice: Patricia Barres, Victoria Bell, Kyung-Soo Do, Zachary Estes, Yevgeniya Goldvarg, Fabien Savary, Lisa Torreano, and Yingrui Yang. Thanks also to Jonathan St. B. T. Evans, Monica Bucciarelli, Vittorio Girotto, and three anonymous reviewers.

© 2006 Psychology Press Ltd

http://www.psypress.com/tar DOI: 10.1080/13546780500264545

To know what is true, individuals sometimes have to think about what is false. Unfortunately, they do not often follow this precept. When they reason, they usually consider only what is true. For example, consider the following testimony in a trial (see Harr, 1995, p. 361). An expert witness conceded the following two points: If the trichloroethylene (TCE) came from the river, then it would be in the riverbed. TCE wasn't in the riverbed. The lawyer, Jan Schlictmann, pointed out that this testimony was consistent with the proposition that TCE did not come from the river. He failed to grasp that this conclusion, in fact, follows *necessarily* from the testimony. One way to draw this inference, however, is to think about the possibilities in which the antecedent of the conditional—TCE came from the river—is false.

According to the theory of mental models (Johnson-Laird & Byrne, 1991), individuals reason from understanding of the premises and their general knowledge. They use this information to construct mental models of the relevant possibilities. For example, given a premise of a form, such as:

There is not an A or else there is a B.

they construct two *mental* models of the two alternative possibilities compatible with this assertion, an exclusive disjunction:

¬A
B

where "¬" denotes negation. These models represent the clauses in the premises only when they are true in a possibility. In contrast, *fully explicit* models represent clauses in models both when the clauses are true and when they are false. The fully explicit models of the exclusive disjunction are accordingly:

¬A ¬B
A B

An equivalent description of these possibilities is:

There isn't an A if and only if there isn't a B.

Individuals tend not to notice this equivalence, presumably because they tend not to construct fully explicit models.

Individuals normally appear to reason on the basis of mental models. But, they can sometimes flesh out these models to make them fully explicit. This fleshing out does occur in conditional reasoning. A conditional of

the form, *If A then B*, such as the one in the TCE example, has the mental models:

A B
...

The first model represents the possibility in which the antecedent, *A*, holds and so the consequent, *B*, also holds. The second model (shown as an ellipsis) has no explicit content, but stands in for the possibilities in which the antecedent of the conditional is false. The mental models suffice for drawing an inference of the form:

If A then B.
A.
Therefore, B.

The categorical premise, *A*, eliminates the implicit model to leave only the first model, which yields the conclusion, *B*. The TCE inference, however, has premises of the form:

If A then B.
Not B.

The categorical premise, *not B*, eliminates the explicit mental model to leave only the implicit model. Because it has no content, it yields no conclusion. Hence, the theory predicts that reasoners should tend to respond, "nothing follows", which is indeed a frequent response to inferences of this form (see Evans, Newstead, & Byrne, 1993). Yet, some reasoners do succeed in drawing the valid conclusion, *not A*. One way in which they can do so depends on remembering that the implicit model represents the possibilities in which the antecedent of the conditional is false. They can use this information to flesh out the models of the conditional into fully explicit ones:

A B
¬A B
¬A ¬B

The categorical premise, *not B*, is consistent with only the third of these models, which yields the conclusion, *not A*. A further source of difficulty may be the need to work opposite to the "directional" bias built into the interpretation of conditionals (Evans, 1993). That is, reasoners have to work backwards from the negation of the consequent to the negation of the antecedent.

What can improve performance, however, is the modulation of a conditional's interpretation by semantic or general knowledge. Such knowledge can add spatial, temporal, or other relations between the antecedent and the consequent, and it can make the model required for the previous inference more salient. For example, the premises:

If Bill is in Rio de Janeiro then he is in Brazil.

Bill is not in Brazil.

readily yield the conclusion:

Bill is not in Rio de Janeiro.

Individuals know that Rio de Janeiro is in Brazil, and so if Bill is not in Brazil then he cannot be in Rio (see Johnson-Laird and Byrne, 2002, for empirical corroboration of this prediction).

In summary, individuals normally reason from mental models, which are constructed according to the principle of *truth*, i.e., a model of a possibility represents clauses in the premises only when these clauses are true in the possibility. The principle applies both to affirmative and to negative clauses, and it does so even when information about falsity is required for a valid inference. The advantage of the principle of truth is that it reduces the load on the processing capacity of working memory. But, the principle has an unexpected disadvantage. It can lead reasoners into systematic and compelling errors. For example, in a study of probabilistic reasoning, Johnson-Laird and Savary (1996) gave their participants the following problem:

Only one of the following assertions is true (about a hand of cards):

If there is a king in the hand then there is an ace.

If there isn't a king in the hand then there is an ace.

Which is more likely to be in the hand: the king or the ace?

Most of participants (79%) responded that the ace was more likely to be in the hand than the king. The principle of truth predicts this response. Individuals consider the case in which the first conditional is true, which yields the mental models:

King Ace

They consider the case in which the second conditional is true, which yields the mental models:

¬King Ace

...

The ace occurs in both sets of models, whereas the king occurs only in the first set, and so they respond that the ace is more likely to be in the hand than the king. What they overlook is that when one conditional is true, the other conditional is false. Individuals consider that a conditional is false when its antecedent is true and its consequent is false. The rubric "only one of the following assertions is true" conveys an exclusive disjunction. Hence, either *if king then ace* is true and *if not king then ace* is false, or *if king then ace* is false and *if not king then ace* is true, it follows that the premises yield two fully explicit models:

King ¬ Ace (the first conditional is false and the second conditional is true)
¬King ¬ Ace (the first conditional is true and the second conditional is false)

The ace is therefore not only less likely than the king, it is also impossible. Only 13% of the participants made this correct response. The fully explicit models of an exclusive disjunction of a biconditional interpretation of the two conditionals, i.e., *if and only if king then ace* and *If and only if no king then ace*, yield all possible contingencies:

King Ace
King ¬Ace
¬King Ace
¬King ¬Ace

The two cards are then equiprobable (and 8% of the participants made this response, although whether they did so as a result of this interpretation is not known). The participants responded reliably better with control problems, which yield the correct response even if individuals rely on mental models, i.e., models that fail to represent what is false. The following problem is a control:

Only one of the following assertions is true (about a hand of cards):
If there is a king, then there is an ace.
If there is a king, then there is not an ace.

The majority of participants responded correctly that the king was more likely to be in the hand than the ace.

Why do individuals succumb to the illusory problems? Formal rule theories (e.g., Braine & O'Brien, 1998; Rips, 1994) postulate that reasoners rely on a set of formal rules of inference, which they use to try to prove conclusions. But, these theories as yet provide no account of the illusory inferences. One difficulty is that the theories rely on logically correct formal rules, and it is difficult to derive systematically invalid conclusions from them. The introduction of invalid rules, however, would seem to be a recipe for wholesale irrationality.

Various theorists have postulated that reasoning depends on dual processes (see, e.g., Johnson-Laird, 1983; Schroyens, Schaeken, & Handley, 2003; Sloman, 1996; Stanovich, 1999). Evans and Over (e.g., 1996) have proposed an influential theory of this sort. They distinguish between rapid, involuntary, implicit reasoning (System 1) and slow, voluntary, explicit reasoning (System 2). System 1 provides the information needed for System 2, which is typically called on for testing hypotheses, solving novel problems, and drawing complex inferences. The focus on truth arises in System 1 (during the comprehension of premises, see Johnson-Laird, 1983). Hence, a potential antidote to illusory inferences should be to engage System 2 in an organised attempt to envisage what is false. Such efforts should lead to the representation of clauses that are false in models, and so they should guide reasoners to the correct responses.

The present experiments were designed to test this prediction. Experiment 1 called for the participants to envisage situations in which assertions were false and to write them down. Experiment 2 used the same procedure, but in addition the participants had to consider the implications of the falsity of assertions. This experiment used the familiar context of a game of checkers. It also examined whether the falsification procedure transferred to a new problem in which the procedure was not explicitly elicited.

EXPERIMENT 1

Individuals succumb to illusory inferences according to the model theory because they fail to think about what is false. If they had to think about these cases, then the theory predicts that they should be less susceptible to illusions. Such a procedure does not provide all the information necessary to make correct responses: Individuals still need to bear in mind what is true too, and to integrate the two sorts of information. Nevertheless, the procedure should help them to make more accurate responses to problems, such as:

> One of the following assertions is true and one of them is false (about your hand of cards):
>
> If there is a king then there is an ace.
>
> If there is not a king then there is an ace.
>
> Which is more likely to be in your hand, the king or the ace?

Method

Participants. Twenty four Princeton University undergraduates, (17 females, 7 males; age range 18–21 years) with no previous training in logic were paid $5.00 to participate in the experiment, which lasted approximately 35 minutes.

Design. The participants acted as their own controls and carried out two blocks of problems. In one block, they had to envisage the circumstances in which each conditional was false (the "falsification" condition). In the other block, they were not instructed to envisage these circumstances (the "no falsification" condition). The participants were randomly assigned to one of two groups in order to counterbalance the order of the two blocks.

Materials. Each problem had the initial rubric, "One of the following assertions is true and one of them is false". We devised six illusory problems: three were based on conditionals, such as the example shown above, and three were based on biconditionals, such as:

One of the following assertions is true and one of them is false:

There is not an ace if and only if there is king.

There is a king.

The mental models of these premises are respectively:

¬Ace King

...

and:

King

Hence, individuals should tend to infer erroneously that the king is more likely than the ace. The fully explicit models of the premises as a whole are as follows:

Ace ¬King
Ace King

and so the correct response is that the ace is more likely than the king. Table 1 presents an example of an illusory problem and its control, each with their mental models and fully explicit models, which were used in the experiment. (See Appendix A for the full list of inferences.) The other instances of illusions based on conditionals or biconditionals were formed by the introduction of additional negations without changing the fundamental logic. Each illusory problem had a matching control problem, which was created by negating one clause in the consequent of the second premise. The aim of these varied problems was to prevent the participants from learning a stereotyped response, and we ensured that the occurrence of negations was matched over the control and illusory problems (because negation is a well-known cause of difficulty, see, e.g., Clark and Chase, 1973; Wason, 1959). The 12 problems were assembled into two blocks, a falsification block and a no falsification block, with each

TABLE 1
Examples of an illusory and a control problem in Experiment 1

Illusory problem	Of the following statements, one is true and one is false.
	If there is a king, then there is not an ace.
	If there is not a king, then there is not an ace.
	Which is more likely, the king or the ace.
Mental models	King ¬ Ace
	¬ King ¬ Ace
Fully explicit models	¬ King Ace
	King Ace
The mental models yield the illusory response: King, but the correct response from the fully explicit models is: Ace.	
Control problem	Of the following statements, one is true and one is false:
	If there is a king, then there is an ace.
	If there is a king, then there is not an ace.
	Which is more likely, the king or the ace?
Mental models	King Ace
	King ¬Ace
Fully explicit models:	King Ace
	King ¬Ace
Both the mental models and the fully explicit models yield the correct response: King.	

block containing three control and three illusory problems. Half of the participants received booklets where the falsification block was presented first; the other half received the no falsification block first. The order in which the problems were presented was counterbalanced within each group.

Procedure. The participants were tested individually in a quiet room. For each problem in the falsification block, they carried out a procedure in which they wrote down what would falsify each of the two premises. For example, they read the following text, which states the first premise:

> "If there is a king in your hand, then there is an ace.
> What can you say to make this a false statement? In other words, under what conditions would this be false?"

They then wrote down their answer and carried out the same task for the second premise. The next page of the booklet stated the conditions in which each assertion would be false:

> "Below the two assertions are paired in a problem. One of them is true, and one of them is false. If the first one is false, there is a king but no ace. If the second one is false, then there is no king and no ace."

Finally, they carried out the problem:

"Of the following assertions, one is true and one is false:

If there is a king in your hand, then there is an ace.

If there is not a king, then there is an ace.

Which is more likely to be in your hand, a king or an ace? ——"

In the no falsification condition, the participants proceeded directly to each problem. The participants were allowed to use paper and pencil in tackling the problems.

Results

For both our experiments, we rank ordered the participants' data and analysed main effects and interactions using Mann-Whitney tests for between-participants comparisons and Wilcoxon tests for within-participant comparisons, correcting for ties when $N > 15$ for Wilcoxon tests, and $N > 10$ for Mann-Whitney tests (Siegel & Castellan, 1988). Because the prediction that falsification should increase correct answers is directional, all p-levels are one-tailed. Table 2 presents the percentages of correct responses. Overall, the participants were more accurate in responding to the control problems (69% correct) than to the illusory problems (31% correct; Wilcoxon's $T^+ = 203$, $n = 20$, $z = 3.93$, $p < .00001$). The participants soon converged on the correct falsifications of the premises, even though they received no feedback. They were 79% correct from the third problem onwards. As predicted, the problems in the falsification condition were answered more accurately (53% correct) than those in the no falsification condition (49% correct). The difference was small, but reliable (Wilcoxon's $T^+ = 98$, $n = 18$, $p < .02$). The participants who carried out the falsification in the first block of trials were more accurate than those who did not carry out falsification in the first block of trials (56% vs. 44% correct;

TABLE 2
The percentages of correct responses to the problems in Experiment 1. The only difference between groups 1 and 2 was in the order of the two blocks of trials: the falsification block and the no falsification block

	Group 1—No falsification first		*Group 2—Falsification first*		
	No falsification	*Falsification*	*Falsification*	*No falsification*	*Overall*
Illusory problems	14	42	28	39	31
Control problems	58	64	72	83	69

Mann-Whitney $W_x = 184.5$, $n = 24$, $z = 2.00$, $p < .02$). Table 2 shows that, as the theory predicted, performance improved for the illusory problems (Wilcoxon's $T^- = 38.5$, $n = 9$, $z = 4.78$, $p < .00001$). Falsification marginally influenced performance on control problems (Wilcoxon's $T^- = 29$, $n = 9$, $z = 1.27$, $p < .10$). To test whether the falsification of premises had a greater effect on illusory than control problems, interaction weights were applied to the number of correct responses for control and illusory problems, which were summed for each group. Falsification had a marginally greater influence on illusory than on control problems (Wilcoxon's $T^- = 34.5$, $n = 9$, $z = 1.52$, $p = .06$).

To ensure that practice effects had no effect on our analysis, we compared the accuracy of illusory problems in the first block of the group who received falsification first against the accuracy of the illusory problems in the first block of the group who did not receive falsification first. We also compared the accuracy of control problems across the first blocks of the two groups. Despite the reduction in data points, we found a pattern similar to that found in the analysis of the full dataset, namely, falsification marginally influenced accuracy of answers to the illusory inferences (Mann-Whitney's $W_x = 170.5$, $N = 24$, $z = 1.34$, $p < .09$), and to the control inferences ($W_x = 175$, $N = 24$, $z = 1.56$, $p < .06$). However, falsifying had a greater effect on the illusory problems than on the controls ($U_x = 132$, $N = 24$, $z = 3.44$, $p < .0003$).

EXPERIMENT 2

Reasoners need to be able not only to generate false instances, but also to apply the results to the problem at hand. To make the consequences of falsification more apparent to individuals, we used a procedure that was familiar, that readily yielded the false instances of compound assertions, such as conditionals, and that made them vivid and memorable (cf. Johnson-Laird, Legrenzi, & Legrenzi, 1972; Wason & Green, 1984). In the game of checkers, players move pieces and can see the outcomes of their moves. In Experiment 2, we presented premises that described the presence and absence of red and black checkers in a region of the board, and we asked the participants to use their knowledge of the game to move a checker in order to falsify a premise.

When you play checkers, your goal is to move pieces diagonally across the board to reach the opposite end of the board. You can move a checker by making it "jump" over an opponent's adjacent checker (see Figure 1), or jump over two of your opponent's checkers (see Figure 2), or merely advance into a vacant position (see Figure 3). When you jump over pieces, they are removed from the board. In our experiment, only checkers in a specified area of the board (its centre two rows) counted as present. For

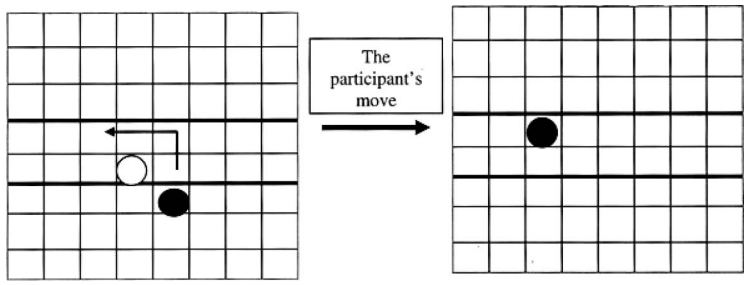

Figure 1. How taking a piece in checkers can represent the falsification of the premise, "If there is a black checker, then there is a red checker". The experimenter says: "Black's turn to move". The participant has the black checker. It jumps over the red checker (shown in white in this figure), as shown in the figure on the left, and the result is to remove the red checker (as shown in the figure on the right). Only those checkers between the two solid lines count, and thus the result is to falsify the premise, because there is a black checker between the two lines but no red checker.

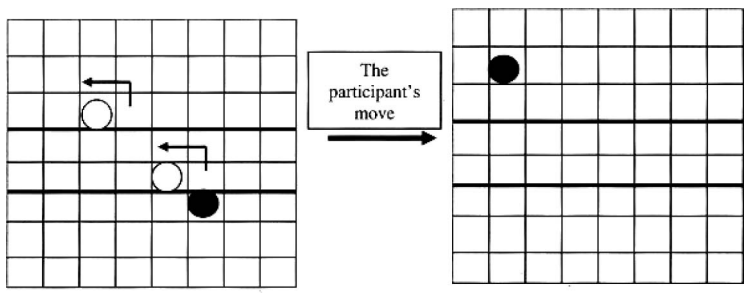

Figure 2. How taking two pieces in checkers can represent the falsification of the premise, "If there is not a black checker, then there is a red checker". The experimenter says: "Black's turn to move". The participant has the black checker (shown on the left side of the figure). It jumps over the red checker (shown in white) that lies between the centre two rows. Because it is jumped by the black checker, the red checker is removed from the centre goal area of the board. Further, the black checker is also removed from the centre area because it jumps over the second red checker. Only those checkers between the two solid lines count, and thus the result is to falsify the premise because there is no black checker and there is no red checker between the two lines.

instance, a red checker in this specified area represented the truth of the assertion: There is a red checker. But, a red checker that was outside this area represented the falsity of the assertion: There is a red checker.

To test whether falsification has effects in general, and not just for the conditionals and biconditionals used in Experiment 1, the present experiment also used conjunctions and inclusive disjunctions. The

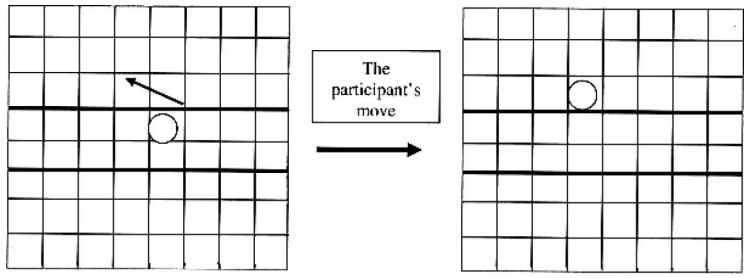

Figure 3. How moving a checker can represent the falsification of the premise, "If there is not a black checker, then there is a red checker". The experimenter says: "Red's turn to move". The participant has the red checker (shown in white in this figure). It advances one square (as shown in the figure on the left), and the result is to move the red checker beyond the centre line (as shown in the figure on the right). Only those checkers between the two solid lines count, and thus the result is to falsify the premise, because there is no black or red checker between the two lines.

following is an example of a control problem in which the premises are conjunctions:

> One of the following statements is false:
>
> The black checker is there and the red checker is there.
>
> The black checker is there and the red checker is not there.
>
> Which is more likely, a black checker or a red checker?

To help participants to envisage false instances of the premises, participants moved the pieces and reported what piece, if any, remained in the middle two rows of the board. To model falsification of the first premise of the problem above, a black checker and a red checker were posed adjacent to each other. Participants used the black checker to jump over the red checker, and the red checker was removed from the board while the black checker stood in the centre area. To help participants consider the implications of one premise being false while the other is true, a column labeled "True" and a column labelled "False" were provided under each premise in the booklet. The experimenter asked, "What is left in the target area?" When the participant answered aloud, e.g., "black and no red", the experimenter replied, "What conditions would make this statement false?" Under a column labelled "False" the participant wrote the conditions that he or she believed made the statement false, e.g., our example participant above wrote "$+B-R$". The participants were then asked to record what conditions would make the premise true under a column marked "True"; e.g., the same participant wrote "$+B\ +R,\ -B\ -R,\ -B\ +R$".

The experimenter then repeated the procedure on the second checkerboard for the second premise: There is a black checker, and there is no red checker. A black checker was already in the middle area, a red checker was just below the middle area, and the participant was told that it was red's turn to move. He or she advanced the red checker by one square, which moved the checker into the middle area. With the red checker now inside the centre two rows, the situation falsified the premise. Under the "True" column for the first premise of the conjunction problem above, one participant wrote " +B +R" , and " +B −R" under the "False" column.[1] After the participant moved the checkers and filled in the columns for both of the premises, the experimenter then asked: "Which coloured checker is more likely to be in the centre area?" Participants could inspect the columns to see which checker was more likely when the first premise is true and the second premise false, and vice versa. The above problem correctly solved is: black, the answer given by the participant referred to in this example.

We predicted that encouraging participants to think of ways a premise can be falsified would improve accuracy on illusory inferences. Different sorts of premise have a different number of situations in which a premise can be false. Premises composed of conditionals and inclusive disjunctions have just one circumstance in which they can be made false, e.g., "If there is a red checker, then there is a black checker" is false only when there is a red checker without a black checker. Biconditionals have two circumstances under which they are false, and conjunctions have three. During the falsification phase of Experiment 2, participants were invited to create *only one* circumstance in which a premise could be false. Therefore, all of the information required for a correct answer to a biconditional or a conjunction would not be recorded under their "False" columns. Unless participants are able to generate further false instances on their own accord, they could be less likely to answer biconditionals and conjunctions correctly.

Method

Participants. We tested 16 Princeton University undergraduates, (10 females, 6 males; age range 17–20 years). They received either course credit or a payment of $6. They were familiar with the game of checkers, but had no previous training in logic.

[1]Although there are two other situations in which the conjunction could be false, participants did not normally provide all three. We predicted that the omissions of the additional false models would not influence answers to control problems but would negatively affect answers to the illusory problems.

Design. The participants were randomly assigned to one of two groups and carried out two blocks of problems. The falsification group tackled the inferences in the first block of trials without any falsification instruction, but in the second block they falsified premises with the aid of the checkers manipulation. A control group solved the problems in both blocks without the falsification aid.

Materials. The problems were presented in booklets, and the cover page of all of the booklets explained that for each problem, one statement was true and one statement was false and that it was very important to pay attention to the fact that one of the statements is false. Within the booklet, each problem was preceded by the statement: "One of the following statements is false". Two conditionals, two inclusive disjunctions, two biconditionals, and two conjunctions were presented in each block. One of each type was an illusion and the other was a control. One of the illusory conditionals is:

If the black checker is there, then the red checker is there.

If the black checker is not there, then the red checker is there.

To avoid any order effects, two versions of each booklet were created in which the problems in one were presented in reverse order in the other. An illusory conditional at the end of the experiment was the same in each booklet and was presented without the checkers manipulation. This problem was:

One of the following statements is false:

If the black checker is there, then the red checker is not there

If the black checker is not there, then the red checker is not there,

It was included to test whether participants in the falsification group could transfer any knowledge that they may have gained to a novel problem.

Procedure. In the first block of both groups, the experimenter read the problems aloud from the booklet. Participants were told that they could produce their answer by thinking silently or aloud. Space was provided under each problem if they wished to generate their answers with a pencil and paper. The participants had unlimited time to answer, but rarely exceeded five minutes. After answering each problem, participants circled a number from 0 to 7 to rate its difficulty. The difficulty scale was included to investigate whether performance on illusory inferences is poor because they are perceived to be more difficult than the control problems.

The second block of the nonfalsification group proceeded in the same way as the first block. Before they were presented with the problems, the participants in the falsification group were reminded of the three ways checkers can be moved. The experimenter then put checkers on two separate checkerboards. Each checkerboard was used to falsify one premise. The experimenter read the first premise to the participant, e.g., "If there is a black checker, then there is red checker". With one red and one black checker on the checkerboard situated diagonally from each other, the experimenter said, "It's black's turn to move". The participant has the black checker and uses it to jump over the red checker (see Figure 1), and the result is to remove the red checker from the centre two rows of the checkerboard. With only the black checker now in the centre two rows, the situation falsifies the conditional: If there is a black checker then there is a red checker. The experimenter asked, "What is left in the target area?" When the participant answered aloud, e.g., "a black checker and no red checker", using the participant's response the experimenter replied, e.g., "If there is a red checker and no black checker in the target area, what would make this statement false?" Under a column labelled "False" the participant wrote the conditions that he or she believed made the statement false, e.g., our example participant above wrote "+B −R". The participant was then asked to record what conditions would make the premise true under a column marked "True"; e.g., the same participant wrote "+B +R, −B −R, −B +R". The experimenter then repeated the procedure on the second checkerboard for the second premise: If there is no black checker, then there is a red checker. A red checker was inside the middle area (Figure 3), and the participant was told that it was red's turn to move. He or she advanced the red checker by one square, which moved the checker outside the middle area. With the red checker now outside the centre two rows, the situation falsified the conditional.

The experimenter asked, "What is left in the target area?" When the participant answered aloud, e.g., "no black and no red", using the participant's response the experimenter replied, e.g., "If there is not black and not a red checker in the target area, which one of the two statements is false?" Under a column labelled "False", our example participant wrote "−B −R". The participant was then asked to record the conditions they thought would make the premise true under a column marked "True". An answer frequently given was "no black, red". Some participants, such as our example participant, recorded all three instances under which the situation is true, "+B +R, −B −R, −B +R".

Finally, the experimenter said: "Since one statement is true and the other is false, which piece is more likely to be in the middle area?" The participants studied their columns that represented true and false conditions of the premises and wrote their answers in the booklet.

After the participants wrote which checker they thought was more likely for a particular problem, they indicated the difficulty they had solving the problem on the 7-point scale.

Results

The control problems were much easier than the illusory problems (68% vs. 29% correct; Wilcoxon's $T^- = 103.5$, $n = 14$, $p < .0002$); see Table 3. The participants overall made 45% correct responses in the first block of trials and 53% correct responses in the second block. The difference was not reliable ($T^+ = 39$, $n = 15$, $p > .51$); practice did not have an overall effect. Similarly, there was no reliable difference between the two groups (Mann-Whitney $U_A = 42$, $z = 1.00$, $p = .16$). However, the falsification group did show a reliable improvement in responding to the illusory problems on the introduction of the falsification procedure in the second block, and this difference was reliable, with lower difference scores between control and illusory inferences occurring in the block in which falsification instruction was presented (control group block 1 d = 16, block 2 d = 13; falsification group block 1 d = 20; block 2 d = 6; $W_x = 309.5$, $z = 1.75$, $p < .05$).

Data from the different connectives were analysed separately and are reported below.

Connective type

Inclusive disjunctions. The theory predicted that the falsification manipulation would improve performance on inclusive disjunctions because they are falsified by only one model. Performance improved in the falsification

TABLE 3
Percentages of correctly answered illusory and control problems in the no falsification and falsification groups of Experiment 2

	No falsification group				*Falsification group*			
	No falsification		*No falsification*		*No falsification*		*Falsification*	
Connective type	*Illusion*	*Control*	*Illusion*	*Control*	*Illusion*	*Control*	*Illusion*	*Control*
Inclusive disjunctions	25	88	0	63	13	88	63	25
Conditionals	13	63	50	63	13	88	75	88
Biconditionals	25	38	38	50	13	25	63	63
Conjunctions	13	88	13	88	25	100	25	75
Overall	19	69	25	66	16	75	57	63

group (12.5% vs. 62.5%), $T^- = 21$, $z = 2.25$, $p < .02$, but not in the control group (25% vs. 0%), $T^- = 1.5$, $z = .00$, n.s.

Conditionals. Performance on illusory inferences improved in the second, falsification, block of the falsification group (12.5% vs. 75%), $T^- = 15$, $z = 2.24$, $p = .025$ However, while nonsignificant, correct answers also increased in the second block of the no falsification group (12.5% vs. 50%), suggesting an additive role of practice for the conditionals.

Biconditionals. Surprisingly, falsification yielded better performance for the biconditionals (12.5% vs. 62.5%), $T^- = 15$, $z = 2.24$, $p < .03$. No practice effect was seen in the nonfalsification group (25% vs. 37.5%). Apparently, participants were able to generalise the strategy and think of an additional circumstance under which the premise would be false. Seven of the eight participants who received the falsification block wrote models for the falsification of both premises of the illusory biconditional.

Conjunctions. Conjunctions were not predicted to be helped by the checkers manipulation because answering them correctly requires three models of what is false. Accuracy in the falsification group was similar to that in the control group (25% vs. 12.5%).

Transfer problem. People in the falsification group were more accurate than controls when solving the transfer problem (50% vs. 0%), Fisher Yates exact test, $p < .04$. Learning how to falsify aided reasoning when instruction was not provided.

Difficulty ratings. In addition to answers to the inferences, we also collected ratings of how difficult participants perceived each inference to be. There appears to be very little consistency between improvement and perceived difficulty. Higher difficulty ratings were reported when people generated more correct answers to the conditionals and biconditionals, but difficulty ratings did not change with the inclusive disjunctions, even though accuracy did improve. Additionally, there were reports of decreased difficulty when people did not receive falsification instruction for the conjunctions and accuracy was low. Overall difficulty ratings for the controls and the illusions did not differ (2.87 controls vs. 2.92 illusions).

DISCUSSION

People often succumb to illusory inferences: They draw conclusions that seem plausible but that are invalid. They are invalid in terms of the judgements that the individuals themselves make in simpler cases. For example, they judge that a conditional is false when its antecedent is true and its consequent is false—a fact corroborated in Experiment 1. But, when they are told that one conditional is true and another conditional is false, they think about the truth of the first conditional, and then about the truth

of the second conditional. And they fail to think about the concurrent falsity of the other conditional, almost as though it had ceased to exist. The model theory predicts the illusions on the grounds that individuals tend to represent only what is true, and that within a model they tend to represent clauses in the premises only when they are true. This failure to represent what is false yields for certain, but not all, premises the prediction that illusory inferences should occur. The theory therefore makes the further prediction that if individuals can be induced to think about what is false—without being swamped by too much information—then the performance with the illusions should improve.

Illusory inferences occur in modal deductions about possibilities (Goldvarg & Johnson-Laird, 2000), and these investigators found one way in which to reduce them. The participants had to check whether their putative conclusions were consistent with the given constraint that only one of the premises was true. Yang and Johnson-Laird (2000) observed illusions in deductions based on quantifiers such as "all" and "some". They were able to reduce the illusions by instructing participants to think explicitly about the consequences of the falsity of a premise.

In contrast, our experiments examined illusions in *probabilistic* reasoning, and corroborated their occurrence. The experiments also corroborated the prediction that the illusions should be reduced when individuals thought about the cases in which the premises were false. In Experiment 1, a typical problem was:

> One of the following assertions is true and one of them is false:
>
> There is not an ace in your hand if and only if there is a king.
>
> There is a king in your hand.
>
> Which is more likely to be in your hand, a king or an ace?

The model theory predicts that individuals who rely on mental models should respond: the king; but the correct answer is: the ace. In one block of trials, the participants had to write down what would falsify the first premise, and then what would falsify the second premise. This procedure produced a robust reduction in the illusions. In the group that received the instruction to falsify in the second block of trials, performance was 200% better. Yet, it still remained below ceiling.

Experiment 2 also required the participants to falsify the premises, but in a more vivid way. They carried out a move in a checkers game that represented the falsification of a premise. Given, say, the premise:

If there is a red checker, then there is a black checker [in a designated area of the board] they moved the red checker so that it jumped over the black checker, which was then removed from the board. The result accordingly

represented a case that falsified the premise, because there was a red checker but no black checker. We examined a variety of logically distinct premises, but once again the effect of falsification was to reduce the illusions. Falsification increased accuracy rates for the illusory inclusive disjunctions and biconditionals but, as predicted, not in conjunctions. Although improvement with the conditionals was to some extent a result of practice, the ameliorative effect of falsification on conditionals was seen in the final trial, where the strategy transferred unbidden to circumstances in which there was no instruction to falsify the premises.

Naïve reasoners appear to rely on mental models that represent only what is true. This phenomenon is consistent with a dual process theory of reasoning (Evans & Over, 1996; Stanovich, 1999): they are relying on a System 1 process. But, the instruction to construct a falsifying instance of a premise is likely to call for a more deliberate process of the sort carried out by System 2. In other words, the normal construction of mental models of discourse is guided by the principle of truth, and corresponds to System 1 processing. But, the construction of fully explicit models, which also represent what is false in possibilities, calls for System 2 processing. It stretches the processing capacity of working memory to the limit, and is accordingly a slower and more deliberate process. The representation of what is true often suffices to reach the right answer in both life and the laboratory. But, there are occasions in which it can be critical to represent what is false.

Wason's selection task (Wason & Johnson-Laird, 1972) is one well-known laboratory task inspired by how scientists test hypotheses. A real-life example in which falsification was also critical is the 19th century physician Ignaz Semmelweis's investigation into childbed fever (cf. Lipton, 1991), a mortal illness that many women suffered after giving birth in the main obstetric hospital in Vienna. There were several hypotheses about the cause of the fever. One hypothesis was that the presence of a priest administering last rites in the ward had a psychological influence on the women that made them susceptible to the disease. To show that this hypothesis was false, Semmelweis removed the priest from the ward, but the mortality rate remained high. Because the women were examined post-partum by medical students who had handled cadavers before their visits to the maternity ward, Semmelweiss hypothesised that "cadaver material" was the culprit. His cure called for the falsification of the premise: If there is childbed fever, then the patient's examiner must have previously handled cadavers. He made sure that the students did not handle cadavers prior to their obstetric examinations. In this way, he arrived at the true cause of the deaths. There was a dramatic decrease in mortality rates.

Falsification may also be important in perception of individuals. A common error is to seek evidence that confirms our attitudes and beliefs about others. This procedure is a recipe for the maintenance of stereotypes.

One way to circumvent a stereotype is to seek *disconfirming* evidence. Thus, to improve Ophelia's reasoning about love, Hamlet advised her: ''Doubt truth to be a liar''. She should have taken his advice.

Manuscript received 4 April 2003
Revised manuscript received 4 July 2005

REFERENCES

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

Clark, H. H., & Chase, W. (1973). On the process of comparing sentences against pictures. *Cognitive Psychology*, *3*, 472–517.

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., Newstead, S. E., & Byrne, R. M. J. (1993). *Human reasoning: The psychology of deduction*. Hove, UK: Lawrence Erlbaum Associates Ltd.

Evans, J. St. B. T., & Over, D. E. (1996). *Rationality and reasoning*. Hove, UK: Lawrence Erlbaum Associates Ltd.

Goldvarg, Y., & Johnson-Laird, P. N. (2000). Illusions in modal reasoning. *Memory and Cognition*, *28*, 282–294.

Harr, J. (1995). *A civil action*. New York: Random House.

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

Johnson-Laird, P. N., & Byrne, R. (1991). *Deduction*. Hove, UK: Lawrence Erlbaum Associates Ltd.

Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review*, *109*, 646–678.

Johnson-Laird, P. N., Legrenzi, P., & Legrenzi, M. (1972). Reasoning and a sense of reality. *British Journal of Psychology*, *63*, 395–400.

Johnson-Laird, P. N., & Savary, F. (1996). Illusory inferences about probabilities. *Acta Psychologica*, *93*, 69–90.

Johnson-Laird, P. N., & Savary, F., (1999). Illusory inferences: A novel class of erroneous deductions. *Cognition*, *71*, 191–229.

Lipton, P. (1991). *Inference to the best explanation*. New York: Routledge.

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

Schroyens, W., Schaeken, W., & Handley, S. (2003). In search of counter-examples: Deductive rationality in human reasoning. *Quarterly Journal of Experimental Psychology*, *56A*, 1129–1145.

Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). McGraw-Hill: New York.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *199*, 3–22.

Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Wason, P. C. (1959). The processing of positive and negative information. *Quarterly Journal of Experimental Psychology*, *11*, 97–107.

Wason, P. C., & Green, D. W. (1984). Reasoning and mental representation. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *36*, 597–610.

Wason, P. C., & Johnson-Laird, P. N. (1972). *The psychology of reasoning: Structure and content*. Cambridge, MA: Harvard University Press.

Yang, Y., & Johnson-Laird, P. N. (2000). How to eliminate illusions in quantified reasoning. *Memory and Cognition*, *28*, 1050–1059.

APPENDIX

The six illusory and the six control problems from Experiment 1 presented here in abbreviated form*. In each case, the task was to say which was more probable, K or A. The illusory answers are incorrect. *If = If and only if

Illusory problems	*Control problems*
1. If K then A If not K then A (Illusory answer: A)	1. If K then A If K then not A (Correct answer: K)
2. Iff not K then A A (Illusory answer: A)	2. Iff not K then A not A (Correct answer: A)
3. Iff not A then K K (Illusory answer: K)	3. Iff not A then not K K (Correct answer: K)
4. Iff K, then not A K (Illusory answer: K)	4. Iff K then not A not K (Correct answer: K)
5. If not K then not A If K then not A (Illusory answer: K)	5. If not K then not A If not K then A (Correct answer: A)
6. If K then not A If not K then not A (Illusory answer: K)	6. If not K then A If not K then not A (Correct answer: A)