Memory & Cognition
2000, 28 (6), 1050-1059

How to eliminate illusions in quantified reasoning

YINGRUI YANG and P. N. JOHNSON-LAIRD
Princeton University, Princeton, New Jersey

The mental model theory postulates that reasoners build models of situations described in premises. These models normally make explicit only what is true according to the premises. The theory has an unexpected consequence. It predicts the existence of *illusions* in inferences: Certain inferences should have compelling but erroneous conclusions. Previous studies have corroborated the existence of such illusions. The present study reports the first effective antidote to them. For example, most people incorrectly answer "yes" to the following problem: *Only one of the following statements is true . . . /At least some of the plastic beads are not red./None of the plastic beads are red./Is it possible that none of the red beads are plastic?* In two experiments, we progressively eliminated this fallacy and others by using instructions designed to overcome the bias toward truth. The difference between the illusory and the control problems disappeared when the participants were instructed to work out both the case in which the first premise was true and the second premise was false and the case in which the second premise was true and the first premise was false.

The ability to reason is central to human cognition. Although some theorists argue that reasoning depends on knowledge or experience, people can make inferences about the unknown. Consider, for instance, the following inference:

All these sonatas are tonal pieces.

All tonal pieces have harmonic relations that can be handled by a context-sensitive grammar.

∴ All these sonatas have harmonic relations that can be handled by a context-sensitive grammar.

Even if one knows nothing about the relevant set of sonatas or about context-sensitive grammars, one can still grasp the validity of the inference. The conclusion resulting from the inference must be true given that the premises are true. The ability to make valid inferences about abstract matters is presumably a precursor to the acquisition of logic and mathematics. Yet it is controversial. Some theorists argue that the making of inferences depends on formal rules that are akin to those of a logical calculus and that reasoners construct a chain of inferential steps akin to those of a proof (e.g., Braine, 1998; Rips, 1994). According to such theories, reasoning is a syntactic process: The logical form of the premises is recovered, and then formal rules are applied to the premises in order to derive a proof in a sequence of syntactic steps (Yang, Braine, & O'Brien, 1998). For example, the following rule of inference can be used to make the preceding inference about the sonatas:

All A are B.

All B are C.

∴ All A are C.

An alternative theory postulates that reasoning is a semantic process. According to this theory, reasoners construct mental models of the situations described by the premises, and they test the validity of conclusions by checking whether they are consistent with all the models of the premises (e.g., Johnson-Laird & Byrne, 1991). Thus, the inference about the sonatas can be made from a single model of the premises. Reasoners assume a small but arbitrary number of tokens to stand for the relevant set of sonatas. They tag each token designating a sonata in order to indicate that the piece is tonal and then that it has harmonic relations that can be handled by a context-sensitive grammar. This model supports the conclusion that all the sonatas have harmonic relations that can be handled by a context-sensitive grammar. Reasoners can search for alternative models of the premises that refute this conclusion, but there is no such model and so the conclusion is valid.

The controversy between these two theories is long standing. Each theory can account for certain empirical phenomena, and there are few crucial results that corroborate one theory and refute the other. But, one seemingly innocuous assumption of the model theory has led to a discovery that, as we shall see, may be able to resolve the controversy.

In order to minimize the load on working memory, reasoners represent as little information as possible. Accordingly, a fundamental assumption of the mental model theory is: the principle of *truth*, which states that the mental models of a set of assertions represent only the true pos-

This research was supported by the Xian-Lin Ji Foundation of Peking University. We thank Jonathan Baron, Mike Oaksford, Steven Sloman, and an anonymous referee for their helpful criticisms of an earlier version of the paper. We also thank the members of our laboratory for much advice: Patricia Barres, Monica Bucciarelli, Victoria Bell, Zachary Estes, Yevgeniya Goldvarg, and Mary Newsome. Address correspondence to Y. Yang, Department of Psychology, Green Hall, Princeton University, Princeton, NJ 08544 (e-mail: yingruiy@phoenix.princeton.edu).

sibilities according to the assertions, and each model of a true possibility represents the literal propositions in the premises (affirmative or negative) only when they are true within the possibility. A *literal* is a proposition that contains no sentential connectives and that is either affirmative or negative. Thus, the conjunction:

There is a circle, but there is not a triangle

contains two literals (there is a circle, there is not a triangle).

The principle of truth is subtle, because it applies at two levels. At one level, mental models do not represent those possibilities that are false according to the premises. In the case of the preceding conjunction, for example, reasoners construct a single mental model of the only true possibility:

o ¬ Δ

where "¬" denotes negation. Mental models do not represent the three cases in which the conjunction is false (the presence of a circle and a triangle, the absence of a circle and the presence of a triangle, and the absence of both a circle and a triangle). We explain below how people try to represent the false possibilities when a task calls for them. As the previous example shows, mental models do represent negative assertions provided that they are true. Negative assertions are a well-known cause of difficulty in reasoning (see, e.g., Evans, Newstead, & Byrne, 1993), but this difficulty does not override the principle of truth; people can represent possibilities corresponding to true negative assertions.

The principle of truth applies at a second level, which concerns the representation of the literals in an assertion. Thus, the exclusive disjunction

There is a circle or else there is not a triangle

has two mental models, one for each of its true possibilities, which we represent on separate lines:

o

¬ Δ

As these models illustrate, a literal in an assertion is represented in a possibility only if it is true in that possibility. Hence, the first of these models represents explicitly that there is a circle, but it does not represent that the literal, *there is not a triangle*, is false in this possibility. Likewise, the second model represents explicitly that there is not a triangle, but it does not represent that the literal, *there is a circle*, is false in this possibility. According to this theory, reasoners try to remember what is false, but these "mental footnotes" soon tend to be forgotten, especially when the assertions contain several connectives.

Mental models are sensitive to the sentential connective; they represent only the possibilities that are true depending on the connective. And within these possibilities, they represent a literal proposition (affirmative or negative) in the premises only when it is true within a possibility. If individuals can keep track of the mental footnotes, they can try to flesh out the mental models in order to convert them into *fully explicit* models. This task calls for forming the complement of a set of models. Hence, to form fully explicit models from the preceding mental models of the disjunction, the first model has to represent, in addition, that it is false that there is not a triangle. Individuals have to form the complement of the model representing that there is not a triangle. This task of envisaging what is false can be difficult (see Barres & Johnson-Laird, 1997). In the present case, it yields a model of the proposition that there *is* a triangle. Likewise, the second model has to represent, in addition, that it is false that there is a circle (i.e., that there is not a circle). The resulting fully explicit models are as follows:

o Δ

¬ o ¬ Δ

The original disjunctive assertion is accordingly equivalent to the biconditional

There is a circle if and only if there is a triangle

Hardly anyone grasps this equivalence when they read the disjunction—a failure that is a testament to the use of mental models and to the difficulty of fleshing out models to make them fully explicit.

If people do need to envisage the possibilities that are false according to an assertion, they construct them by forming the *complement* of the fully explicit models. That is, they take the fully explicit models of the true possibilities—for example,

o Δ

¬ o ¬ Δ

—and form the complement of these models:

¬ o Δ

o ¬ Δ

Once again, this task of negating a set of models is difficult (see Barres & Johnson-Laird, 1997).

The principle of truth has an unexpected consequence, which we discovered by accident in the output of a computer program implementing the model theory. Most valid inferences can be made from models that represent only what is true according to the premises, but there are some inferences in which such models should lead reasoners systematically astray. These inferences should accordingly yield systematic fallacies (i.e., invalid conclusions that most individuals infer). These systematic fallacies do occur, and they are often so compelling that they amount to cognitive illusions. For example, a study of inferences about what is possible contained the following premises (Goldvarg & Johnson-Laird, 2000):

One of the following premises is true about a particular hand of cards and one is false:

There is a king in the hand, or there is an ace, or both.

There is a queen in the hand and there is an ace.

These premises were combined on separate trials with questions of different sorts. In one case, the question was as follows:

> Is it possible that there is a queen in the hand and an ace?

The majority of participants wrongly inferred that the answer was "yes" (as predicted by the model theory). They considered the case in which the second premise was true and overlooked that, in this case, the first premise must then be false. The illusions are robust, and perhaps the most compelling example is illustrated by the following problem about a particular hand of cards (Johnson-Laird & Savary, 1999):

> If there is a king in the hand then there is an ace in the hand, *or else* if there is not a king in the hand then there is an ace in the hand.
>
> There is a king in the hand.
>
> What follows?

All the participants concluded that there was an ace in the hand. Mental models yield this conclusion, even though it is wrong. In fact, the sentential connective *or else* means that one of the conditionals is false (or may be false granted an inclusive interpretation). Thus, the first conditional could be false, and, in this case, even though there is a king in the hand, there is no guarantee that there is an ace. The occurrence of illusions has been demonstrated in deductive, modal, and probabilistic reasoning, and the illusions occur with a variety of sentential connectives, including conditionals, disjunctions, and conjunctions (see Goldvarg & Johnson-Laird, 2000; Johnson-Laird & Savary, 1996, 1999).

In a previous study, we demonstrated the occurrence of illusions by using quantified assertions (see Yang & Johnson-Laird, 2000). For example:

> Only one of the following statements is true:
>
> At least some of the plastic beads are not red, or
>
> None of the plastic beads are red.
>
> Is it possible that none of the red beads are plastic?

The model theory predicts that reasoners will construct a mental model of the first premise:

```
p    ¬ r
p    ¬ r
       r
       r
  ...
```

In this case, unlike in the previous diagrams, each row represents a separate individual in the same situation. Thus, there are four beads represented explicitly: "p" denotes plastic, "r" denotes red, "¬" denotes negation, and the ellipsis (the three dots) allows that there may be other sorts of beads. This model is consistent with the possibility that none of the red beads is plastic, so reasoners should tend to respond "yes" to the question. Perhaps, as a reviewer reminded us, this premise leads reasoners to assume that at least some of the plastic beads *are* red, though we used the phrase *at least some* to try to minimize this interpretation. Those who make this assumption, however, will construct the following sort of mental model:

```
p    ¬ r
p    ¬ r
p     r
      r
  ...
```

which is no longer consistent with the possibility that none of the red beads is plastic. However, a mental model of the second premise is

```
[p]  ¬ r
[p]  ¬ r
      [r]
      [r]
  ...
```

—where the square brackets indicate that a set has been exhausted, so that the beads denoted by the ellipsis cannot include any plastic beads or any red beads. This model is certainly consistent with the possibility that none of the red beads is plastic, so reasoners are still likely to respond "yes." Our study showed that, as predicted, most participants responded "yes" (80%). In fact, this response is a fallacy; it is impossible for none of the red beads to be plastic. The fallacy arises, according to the theory, because reasoners fail to take into account that when one premise is true, the other premise is false. When the first premise is true, the second premise is false; that is, some of the plastic beads are red, and so the correct model of this case is

```
p     r
p    ¬ r
  ...
```

Conversely, when the second premise is true, the first premise is false; that is, all of the plastic beads are red, which conflicts with the truth of the second premise, so the result is the empty (or null) model. The only model of the premises is, therefore, the preceding one, and it refutes the possibility that none of the red beads are plastic. The correct answer to the question is accordingly "no."

One normative concern that readers may share with Michael Oaksford (personal communication, January, 1999) is the distinction between language and logic. An assertion of the form, *All X are Y*, in modern logic, makes no claim about the existence of Xs. Hence, as Oaksford

points out, *all ravens are black* and *no ravens are black* are both true in the case in which there are no ravens. To obviate the interpretation in which a universal assertion can be vacuously true, in our experiments we used assertions containing the definite article—that is, assertions akin to *All the ravens are black*. The standard logical analysis of such assertions is that the force of the definite article is either to assert or to presuppose the existence of a set of ravens. Likewise, our instructions to the participants made clear that members existed for all the sets referred to in the premises (see Johnson-Laird & Bara, 1984, for further discussion of this point).

Previous studies have established that illusions are not a result of the participants' ignoring the rubric to the problem (i.e., that only one of the two assertions is true). First, "think aloud" protocols have shown that participants succumb to illusions when they take the premises to be in a disjunction (Johnson-Laird & Savary, 1999). Second, the illusions occur when the rubric is replaced with a sentential connective, such as *or else* (see the previous example). Third, certain control problems would have contradictory premises if participants ignored the rubric, but participants do not treat them as contradictory (Johnson-Laird & Savary, 1999). Fourth, consider the following problem (Goldvarg & Johnson-Laird, 2000):

> Only one of the following premises is true about a particular hand of cards:
>
> There is a king in the hand or there is an ace, or both.
>
> There is a queen in the hand or there is an ace, or both.
>
> There is a jack in the hand or there is a ten, or both.
>
> Is it possible that there is an ace in the hand?

Nearly every participant responded "yes" incorrectly. In a further study, there was a large and reliable improvement in performance when the participants were asked to check whether their conclusions were consistent with the truth of only one of the premises. Yet only 57% of their conclusions were correct. In other words, illusions still occurred on nearly half of the trials despite the remedial instructions, which certainly made clear the nature of the rubric. In sum, illusions are robust and appear to occur because people tend not to represent what is false according to the premises. Illusions are so pernicious that so far no antidote has been able to eliminate them (see Newsome & Johnson-Laird, 1996; Tabossi, Bell, & Johnson-Laird, 1999). They persist over many trials and nothing appears to be able to eliminate them. Hence, it is important to develop an antidote, in part because of its the intrinsic interest, but also because it might illuminate the underlying cause of illusions. Accordingly, our chief aim in the present study was to develop a successful antidote.

EXPERIMENT 1

According to the model theory, illusions arise from a failure to take falsity into account. It follows that instructions designed to inculcate a greater attention to falsity should reduce the tendency to commit the systematic fallacies. Our goal in Experiment 1 was to test the effects of remedial instructions that asked the participants to envisage explicitly the case in which the first premise was true and the second premise was false. The participants acted as their own controls: In the first half of the experiment, they carried out the standard task, and then, in order to maximize the chances of success, they were given the remedial instructions. We examined illusory inferences and also control problems to which the participants should make the correct responses even if they failed to take falsity into account.

Method

Design. The participants carried out four sorts of modal inferences concerning what was possible. The inferences were based on five pairs of indicative premises that each referred to the same two terms and five pairs of related deontic premises. The five pairs of premises were combined on separate trials with different modal conclusions, making a total of 12 different indicative problems and 12 related deontic problems. Each of the indicative problems had the following form:

> Only one of the following statements is true:
>
> Premise 1, or
>
> Premise 2
>
> ---
>
> Is it possible that . . . ?

Each of the deontic problems had the following form:

> You must act to make one, but only one, of the following statements true:
>
> Premise 1, or
>
> Premise 2.
>
> ---
>
> Are you allowed to make . . . ?

The four sorts of inferences were as follows:

1. *Illusions of possibility*, to which the participants should respond "yes" when, in fact, the correct answer is "no." We refer to these problems as "yes/no" problems, an abbreviation that states the predicted answer followed by the correct answer. An example of such a problem has the following form:

> Only one of the following statements is true:
>
> At least some of the A are not B, or
>
> None of the A are B.
>
> Is it possible that none of the B are A?

2. *Controls for these illusions*, to which the participants should respond "yes" correctly ("yes/yes" problems). For example:

> Only one of the following statements is true:
>
> At least some of the A are B.
>
> All the A are B.
>
> Is it possible that some of the B are A?

3. *Illusions of impossibility*, to which the participants should respond "no" when, in fact, the correct answer is "yes" ("no/yes" problems). For example,

Only one of the following statements is true:

At least some of the A are not B.

At least some of the B are not A.

Is it possible that all the A are B?

4. *Controls for these illusions*, to which the participants should respond "no" correctly ("no/no" problems). For example,

Only one of the following statements is true:

All the A are B.

All the B are A.

Is it possible that none of the A are B?

We used 12 problems, 3 of each sort, which were selected from a set of 20 that we had tested in a previous study (Yang & Johnson-Laird, 2000). They consisted of the six illusory problems with the highest error rate and the six control problems with the lowest error rate. Table 1 presents the full set of problems in indicative forms, their mental models, their fully explicit models, and the questions for each pair with the predicted and correct answers. The 24 problems (of which 12 were indicative and 12 deontic) had a different lexical content, and they were presented in one of six different random orders with approximately an equal number of participants receiving each order.

Materials. The problems concerned beads of different colors (blue, red, green, or brown), shapes (square, round, triangular, or rectangular), and materials (wood, plastic, metal, or cement). We chose properties at random to make 24 sets of materials and then assigned them at random to the problems.

Procedure. The participants were tested individually in a quiet room. They were given a booklet containing the initial instructions and a practice problem. The experimenter read the instructions aloud while the participants followed along in the booklet. The instructions stated that the participants' task was to answer a series of questions about the various possibilities regarding information that they would receive about beads. The instructions explained that all the different sorts of beads existed within the hypothetical domain of the experiment. For example:

> All the problems concern the following situation. First, assume that there is a group of children and each of them has a bag. Second, imagine that there are many beads and that the manufacturer puts them in bags. The beads may vary in color, shape, and material.

The instructions also made very clear how to interpret the initial rubric in each problem, which stated that only one of the two fol-

Table 1
The Premises, Their Mental Models and Fully Explicit Models, and the Four Questions and Their Predicted and Correct Answers for Experiment 1, Where "Yes/No" Indicates That the Predicted Answer is "Yes," But the Correct Answer is "No"

Premises and Questions	Mental Models		Fully Explicit Models	
Only one is true: Some A are not B No A are B	a ¬ b a ¬ b b b	[a] ¬ b [a] ¬ b [b] [b]	a ¬ b a ¬ b	
1. Possible no B are A?		Illusion of possibility:		Yes/No
2. Possible all A are B?		Control for "no":		No/No
Only one is true: Some A are B. All A are B.	a b a b	[a] b [a] b	a b a ¬ b	
3. Possible all A are B?		Illusion of possibility:		Yes/No
4. Possible some B are A?		Control for "yes":		Yes/Yes
5. Possible no A are B?		Control for "no":		No/No
Only one is true: Some B are not A. Some A are B.	b ¬ a b ¬ a a a	a b a b	[b] a [b] a	[a] ¬ b [a] ¬ b [b] [b]
6. Possible some A are not B?		Control for "yes":		Yes/Yes
7. Possible no A are B?		Illusion of impossibility:		No/Yes
Only one is true: Some A are not B. Some B are not A.	a ¬ b a ¬ b b b	b ¬ a b ¬ a a a	[b] a [b] a a	[a] b [a] b b
8. Possible no A are B?		Illusion of possibility:		Yes/No
9. Possible some A are B?		Control for "yes":		Yes/Yes
10. Possible all A are B?		Illusion of impossibility:		No/Yes
11. Possible all B are A?		Illusion of impossibility:		No/Yes
Only one is true: All A are B. All B are A.	[a] b [a] b	[b] a [b] a	[a] b b	[b] a [b] a a
12. Possible no A are B?		Control for "no":		No/No

lowing premises was true: "You will notice that every problem contains two statements, but only one of them is true, i.e., one is true and the other is false, though you do not know which of them is true." After the participants had asked questions about the task, they carried out a simple yes/yes problem for practice:

Only one of the following statements is true:

At least some of the brown beads are round, or

All the brown beads are round.

Is it possible that at least some of the brown beads are round?

The experimenter made sure that the participants understood that it was always the case that one premise was true and one premise was false. Once the participants were clear on this point and on the nature of the task, they proceeded to the experiment proper. The first 12 problems were presented in a booklet, with each problem on a separate page. The booklet included all 12 problems (see Table 1) with two versions: Either the even numbered problems were indicative and the odd numbered problems were deontic, or *vice versa*. Each participant had one sort of booklet in the first half of the experiment and the other sort of booklet in the second half of the experiment. Hence, each participant encountered a particular problem only once.

After the participants had completed one booklet of problems, they received the following remedial instructions that were designed to improve their performance:

To solve these problems correctly, you need to do the following: 1) select your response; 2) go back and check whether your response preserves the relationship between the premises, i.e., one of them is still true and the other is still false. For example, suppose you have the following problem:

Only one of the following statements is true:

All of the plastic beads are red, or

Only the plastic beads are red.

Is it possible that at least some of the plastic beads are red?

If you respond: 'Yes' (you believe that at least some of the plastic beads may be red), then go back and check that one of the premises could still be true and the other could still be false. The first premise could be true, that is, all the plastic beads are red. So at lease some of them are red. And the second premise could be false—if, say, the metal beads are also red. Thus, your response is correct: it IS possible that at least some of the plastic beads are red. Please do this checking for every problem, without it you will get many of the problems wrong.

The experimenter made sure that the participants understood these instructions before they proceeded to work on the second booklet of problems.

Participants. Twenty Princeton undergraduates were either paid $6 for their participation or took part in the experiment to fulfill a requirement of their psychology major. They had not received any training in formal logic and had not participated in an experiment on reasoning before.

Results and Discussion

There were no significant differences between the indicative problems (68% correct) and the deontic problems (69% correct), so we decided to pool the results. Table 2 presents the percentages of correct responses for each of the four sorts of inferences, both with and without the remedial instructions. Table 3 presents these percentages for the individual problems. We used Wilcoxon tests, so we report only the values of z and the probabilities. The results confirmed the model theory's predictions.

First, the participants made a greater percentage of accurate responses to the control problems than to the illusory problems ($z = 3.39$, $p < .001$). The effect was particularly marked in the absence of remedial instructions, 93% correct responses to the control problems, 33% correct responses to the illusory problems ($z = 3.82$, $p < .001$).

Second, the remedial instructions led to an overall improvement in accuracy (63% correct without them vs. 72% correct with them; $z = 2.10$, $p < .05$). But, as expected, the effect of the instructions was much greater on the illusory problems than on the control problems ($z = 3.52$, $p < .01$). Indeed, there was a significant improvement in performance with the illusory problems (33% correct without the instructions, 65% correct with the instructions; $z = 3.27$, $p < .01$), but a reliable decline in performance with the control problems (93% correct to 79% correct; $z = -2.31$, $p < .01$). The decline with the controls shows that the improvement with the illusory problems was not merely a practice effect. It takes work to analyze whether a conclusion is consistent with the truth of the first premise and the falsity of the second premise. This work is necessary for the fallacies, but *not* for the control problems, where the neglect of falsity does not lead to error. It may lead to confusions with the controls that would not otherwise occur when reasoners consider only the consequences of truth. A reviewer (Jonathan Baron) wondered whether the remedial instructions had their effect merely because the participants felt under some pressure to change their initial responses as a result of checking them. There may have been such an effect, but the results show that this factor cannot be the whole

Table 2
The Percentages of Correct Responses to the Four Sorts of Problems in Experiment 1 Without and With Remedial Instructions

	Illusions		Controls		Overall	
	Without Remediation	With Remediation	Without Remediation	With Remediation	Without Remediation	With Remediation
Inferences of possibility	25	52	93	73	60	63
Inferences of impossibility	42	78	92	85	67	82
Overall	33	65	93	79	64	72

Table 3
The Percentages of Correct Responses to Each of the Twelve Problems in Experiments 1 and 2 Without and With the Remedial Instructions

		Percentages of Correct Responses			
		Experiment 1		Experiment 2	
Premises and Conclusions	Status of Question	Without	With	Without	With
Only one is true: Some A are not B. No A are B.					
1. Possible that no B are A?	Yes/No	20	35	45	60
2. Possible that all A are B?	No/No	95	80	95	95
Only one is true: Some A are B. All A are B.					
3. Possible that all A are B?	Yes/No	15	65	55	75
4. Possible that some B are A?	Yes/Yes	95	80	85	50
5. Possible that no A are B?	No/No	95	90	95	80
Only one is true: Some B are not A. Some A are B.					
6. Possible that some A are not B?	Yes/Yes	95	80	75	65
7. Possible that no A are B?	No/Yes	35	60	35	50
Only one is true: Some A are not B. Some B are not A.					
8. Possible that no A are B?	Yes/No	45	45	55	75
9. Possible that some A are B?	Yes/Yes	95	60	80	80
10. Possible that all A are B?	No/Yes	40	85	20	85
11. Possible that all B are A?	No/Yes	50	90	25	80
Only one is true: All A are B. All B are A.					
12. Possible that no A are B?	No/No	85	85	90	60

story. If the remedial procedure merely induced a pressure to change an initial response, such changes should not differ between the illusory and control problems. But, in fact, the improvement for the illusory problems was reliably larger than the impairment for the control problems ($z = 2.04$, $p < .05$).

Third, inferences of possibility tend to be more compelling than inferences of impossibility ($z = 2.48$, $p < .05$; see also Goldvarg & Johnson-Laird, 2000; Yang & Johnson-Laird, 2000). Reasoners are more likely to draw a conclusion on the basis of a single model than on the basis of all the models of the premises. Thus, reasoners are more likely to infer that a conclusion is possible than to infer that it is impossible (see Bell & Johnson-Laird, 1998). An alternative possibility is that the illusions of possibility are more compelling because some of them have conclusions that are identical to one of the premises; however, this never occurs for the illusions of impossibility. In fact, the data show that the illusions of possibility are not more powerful when the putative conclusion matches a premise (see also Yang & Johnson-Laird, 2000, for further discussion). Although the results corroborated the model theory's prediction, the antidote failed to eliminate the difference between the illusory and the control problems. Likewise, there was some variation in performance with the individual problems (see Table 3). Therefore, we needed to determine whether the phenomena were robust and, indeed, whether we could devise a more successful antidote.

EXPERIMENT 2

In Experiment 1, we confirmed the existence of illusions that are based on quantified statements. We also showed that performance improved when the participants were instructed to check whether a putative conclusion was consistent with the truth of the first premise and the falsity of the second premise. In Experiment 2, we examined the effects of a more comprehensive antidote. The participants were instructed to check that a putative conclusion was consistent, first, with the truth of the first premise and the falsity of the second premise, and, second, with the truth of the second premise and the falsity of the first premise. These instructions should further reduce the difference between the illusory inferences and the control inferences.

Method

Design and Materials. The design and materials were identical to those of Experiment 1. The participants carried out four sorts of inferences (inferences of possibility and impossibility that were either illusory or controls).

Procedure. The participants were tested individually, and the procedure was almost identical to that of Experiment 1. The major change was in the remedial instructions, which now spelled out explicitly the need to consider two cases:

> To solve these problems correctly, you need to do the following: 1) select your response; 2) go back and check whether your response preserves the relationship between the premises, i.e., one of them is still true, and the other is still false. Remember that when one statement is true, the other statement is false, and that you need to take into account both these facts. For example,
>
> One of the following statements is true and one of them is false:
>
> There is an ace or there is a king, or both;
>
> There is an ace.
>
> Is the above description consistent with the following possibility: There is a king?
>
> You need to consider two possible cases as below, and the questioned conclusion is possible when it is consistent with at least one of the cases.
>
> Case 1. Suppose that the first statement is true, then it follows that the second statement is false, so there is not an ace. It follows from the first statement that there is a king.
>
> Case 2. Suppose that the second statement is true: There is an ace. But it also follows that the first statement is false; i.e., there is not an ace and there is not a king. This contradicts the fact that there is an ace. And so this case is impossible.
>
> Hence, … the first statement must be true and the second statement must be false. And so it follows that there is not an ace, but there is a king.

The participants were instructed to do this check for every problem.

As these instructions show, we also changed the wording of each problem in order to clarify it. For the indicative problems, the questions were posed by the phrase, "Is the description above consistent with the following possibility?" For the deontic problems, the questions were posed by the phrase, "Is the description above consistent with an action that brings about the following possibility?"

Participants. Twenty undergraduates from the same population as before participated in the experiment. As before, they were either paid $6 or participated in order to fulfill a course requirement. They had not received any training in formal logic and had not participated in an experiment on reasoning before.

Results

Table 4 presents the percentages of correct responses for each of the four sorts of inference, both with and without the remedial instructions. Table 3 presents these percentages for the individual problems. We have again collapsed the results for the indicative problems (70% correct) and for the deontic problems (64% correct) because there were no reliable differences between them. Again, we used Wilcoxon tests. The results confirmed the model theory's predictions. First, the participants made a greater percentage of accurate responses to the control problems than to the illusory problems ($z = 3.23$, $p < .01$). The effect was again particularly marked in the absence of the remedial instructions, 87% correct responses to the control problems, but only 39% correct responses to the illusory problems ($z = 3.66$, $p < .001$). Second, the remedial instructions improved accuracy (from 63% correct without them to 71% correct with them; $z = 2.10$, $p < .05$). But, as in Experiment 1, the effect of the instructions was much greater on the illusory problems than on the control problems ($z = 3.41$, $p < .01$). The improvement with the illusions was reliable (from 39% correct to 71% correct; $z = 3.72$, $p < .001$); the decline with control problems was reliable (from 87% correct to 72% correct; $z = 1.99$, $p < .05$). Indeed, there was no reliable difference between the illusions and the controls after the remedial instructions ($z = -0.44$, n.s.).

GENERAL DISCUSSION

The experiments showed that intelligent individuals who have no training in logic succumb to systematic fallacies in quantified reasoning. For example, given the following problem:

> Only one of the following statements is true:
>
> At least some of the plastic beads are not red, or
>
> None of the plastic beads are red.
>
> Is it possible that none of the red beads are plastic?

only 20% of the participants in Experiment 1 answered correctly, "no"; the remaining 80% responded incorrectly, "yes." These responses were predicted on the grounds that individuals fail to cope with what is false. They consider that the putative conclusion is consistent with the truth of either premise, and they fail to take into account that when one premise is true, the other premise is false. Yet, if the first premise is false, then all the plastic beads are red, and therefore some red beads are plastic; if the second premise is false, then some plastic beads are red, and therefore some red beads are plastic. Either

Table 4
The Percentages of Correct Responses to the Four Sorts of Problems in Experiment 2 Without and With Remedial Instructions

	Illusions		Controls		Overall	
	Without Remediation	With Remediation	Without Remediation	With Remediation	Without Remediation	With Remediation
Inferences of possibility	51	70	80	65	66	68
Inferences of impossibility	26	72	93	78	60	75
Overall	39	71	87	72	64	71

way, it is impossible that none of the red beads is plastic. In contrast, the participants coped well with the control problems in which the neglect of falsity did not prevent them from reaching the correct response.

The experiments also developed, for the first time, an effective antidote to illusions. In Experiment 1, we taught the participants to consider the case in which the first premise was true and the second premise was false. This procedure applied to the preceding problem yields the conclusion that some of the plastic beads are red (from the falsity of the second premise), and so it should curb fallacies, as indeed it did in Experiment 1. Yet, the procedure did not close the gap in performance between the illusions and the control problems.

However, Experiment 2 did close the gap. The participants had to analyze two cases. First, they had to analyze the case in which the first premise was true and the second premise was false (Experiment 1). Second, they had to analyze the case in which the second premise was true and the first premise was false. With this procedure, performance on the fallacies and controls converged at around 71% correct, reliably above chance in both cases. This procedure, if it is executed properly, is taxing, so we conclude that the participants from our population had an ability to reach correct conclusions by using the procedure at this level of performance. An ideal antidote, of course, should result in 100% correct performance with both fallacies and controls. Other experiments have explored a variety of antidotes (see Goldvarg & Johnson-Laird, 2000; Tabossi et al., 1999), but, unlike our Experiment 2, none of them eliminated the difference in difficulty between the fallacies and controls. Part of the problem is the difficulty that naive individuals have in thinking about falsity. Barres and Johnson-Laird (1997) have shown that reasoners do not have direct access to the cases in which assertions containing sentential connectives are false, but rather they must infer such cases from their knowledge of cases in which assertions would be true.

One reviewer wondered whether the remedial effect of our instructions might have been a result of asking the participants to check their initial responses. We are skeptical about this possibility, because the illusions are difficult to eliminate, and none of our previous remedial efforts had been successful. In one study, for example, we warned the participants in one condition that some of the inferences were extremely tricky, and we recorded their "think aloud" protocols (see Johnson-Laird & Savary, 1999). The warning led them to check their answers, but had no effect whatsoever on their tendency to succumb to illusions.

Could our results be accounted for by current theories based on formal rules of inference (see, e.g., Braine, 1998; Rips, 1994)? The short answer is that such theories can explain neither illusory inferences nor the effects of our remedial procedures. We could save Rips's theory by exploiting its computational power: It is equivalent to a universal Turing machine, so we could reconstruct within it any computable theory, including the mental model theory itself. Otherwise, rule theories cannot explain illusions because these theories rely solely on valid principles of reasoning, and these valid principles cannot explain systematic invalidity. Similarly, the formal rule theories do not use truth tables or any machinery corresponding to mental models, so they have no way to explain the effectiveness of a procedure in which reasoners consider those possibilities that are false according to the premises. In fact, we know of only one way in which formal rule theorists have tried to account for illusory inferences. Both Luca Bonatti and David O'Brien (personal communications, April 1997) suggested that reasoners use a suppositional strategy designed for conjunctions and misapply it to a disjunction of premises. This hypothesis, as we have shown previously (Yang & Johnson-Laird, 2000), yields the wrong predictions for some control problems. Consider problem 6 in Table 3:

> Only one statement is true:
>
> Some B are not A.
>
> Some A are B.
>
> Is it possible that some A are not B?

The conclusion cannot be validly derived from a supposition of either premise, so reasoners should have responded "no." Yet, most of the participants correctly responded "yes." Likewise, if the misapplied strategy were the correct explanation of illusions, there would be no reason to suppose that our remedial procedures would be effective.

Proponents of rule theories may be tempted to argue that the participants had merely ignored the rubric that only one assertion was true and then reasoned correctly. As we have argued, however, the evidence from previous experiments shows that this assumption is false. When we collect "think aloud" protocols from reasoners, it is clear that they think about the premises as disjunctive alternatives. Likewise, the illusions occur when the rubric is replaced by a sentential connective, such as *or else* (see Johnson-Laird & Savary, 1999). This possibility is still more remote in the present study because we took great pains in the instructions to spell out that only one of the premises was true, and we also worked through an example with the participants in order to explain this point. Another potential criticism is that the illusions are remote from inferences in daily life. But a rubric of the form, "Only one of the following assertions is true," is equivalent to an exclusive disjunction, and a search of the World Wide Web revealed several examples of illusory inferences based on such disjunctions. For instance, a chemistry professor warned his students that

> Either a grade of zero will be recorded if your absence [from class] is not excused, or else if your absence is excused other work you do in the course will count.

The mental models of this assertion yield the two possibilities that presumably he and his students had in mind:

¬ excused	zero-grade
excused	other-work-counts

—but these possibilities are illusory. The fully explicit models of an assertion of the form:

B if not A, or else if A then C

are very different. Indeed, what the professor should have asserted was a conjunction of the two conditionals in order to yield the two possibilities above.

The model theory postulates that individuals normally focus on what is true according to the premises and neglect what is false. It follows that certain illusory inferences should occur. It also follows that any procedure that focuses attention on falsity should improve performance with illusory inferences. Our results corroborate both of these predictions. Of course, the fact that a prescriptive procedure (in which the consequences of falsity are considered) is effective does not in itself establish that the original cause of the error was a neglect of falsity. But, the failure to corroborate our prediction would have overturned this account of illusions. Other studies of illusory inferences also bear out our theory. Thus, Tabossi et al. (1999) obtained an improvement in performance using a rubric that emphasized falsity: "Only one of the following assertions is false." The diversity of illusions appears to have in common one underlying general principle: People have great difficulty in coping with what is false according to the premises. They neglect false possibilities and the falsity of those literal propositions in the premises (affirmative or negative) that are false in a true possibility. Other phenomena can be interpreted in the same light—from the difficulty of *modus tollens* inferences (of the form *If A then B*, *not-B*, *therefore*, *not-A*) to the difficulty of the abstract form of Wason's selection task (see Evans et al., 1993). Even though the model theory ultimately may be overturned by a better theory, such a theory is likely to accommodate the principle of truth. To represent both what is true and what is false according to the premises is equivalent to constructing a truth table, and Osherson (1974–1976) showed that truth tables are an implausible psychological model. A rational compromise is to give up falsity in favor of truth. Truth is more useful than falsity, but the failure to represent falsity exacts its price, in that reasoners may be misled into systematic fallacies.

REFERENCES

Barres, P. E., & Johnson-Laird, P. N. (1997). Why is it hard to imagine what is false? In *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society* (pp. 475-478). Mahwah, NJ: Erlbaum.

Bell, V., & Johnson-Laird, P. N. (1998). A model theory of modal reasoning. *Cognitive Science*, **22**, 25-51.

Braine, M. D. S. (1998). Steps towards a mental predicate logic. In M. D. S. Braine & D. P. O'Brien (Eds.), *Mental logic* (pp. 273-331). Mahwah, NJ: Erlbaum.

Evans, J. St. B. T., Newstead, S. E., & Byrne, R. M. J. (1993). *Human Reasoning: The psychology of deduction*. Hillsdale, NJ: Erlbaum.

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

Johnson-Laird, P. N., & Bara, B. G. (1984). Logical expertise as a cause of error. *Cognition*, **17**, 183-184.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hillsdale, NJ: Erlbaum.

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.

Newsome, M. R., & Johnson-Laird, P. N. (1996). An antidote to illusory inferences. In *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society* (p. 820). Hillsdale, NJ: Erlbaum.

Osherson, D. N. (1974–1976). *Logical abilities in children* (Vols. 1–4). Hillsdale, NJ: Erlbaum.

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

Tabossi, P., Bell, V.A., & Johnson-Laird, P. N. (1999). Mental models in deductive, modal, and probabilistic reasoning. In C. Habel & G. Rickheit (Eds.), *Mental models in discourse processing and reasoning*. Berlin: John Benjamins.

Yang, Y., Braine, M. D. S., & O'Brien, D. P. (1998). Some empirical justifications of the mental predicate logic model. In M. D. S. Braine & D. P. O'Brien (Eds.), *Mental logic* (pp. 333-365). Mahwah, NJ: Erlbaum.

Yang, Y., & Johnson-Laird, P. N. (2000). Illusions in quantified reasoning: How to make the impossible seem possible, and vice versa. *Memory & Cognition*, **28**, 452-465.

(Manuscript received May 17, 1999;
revision accepted for publication October 15, 1999.)