Memory & Cognition
2009, 37 (5), 615-623
doi:10.3758/MC.37.5.615

Disjunctive illusory inferences and how to eliminate them

Sangeet Khemlani and P. N. Johnson-Laird
Princeton University, Princeton, New Jersey

The mental model theory of reasoning postulates that individuals construct mental models of the possibilities in which the premises of an inference hold and that these models represent what is true but not what is false. An unexpected consequence of this assumption is that certain premises should yield systematically invalid inferences. This prediction is unique among current theories of reasoning, because no alternative theory, whether based on formal rules of inference or on probabilistic considerations, predicts these illusory inferences. We report three studies of novel illusory inferences that depend on embedded disjunctions—for example, premises of this sort: *A or else (B or else C)*. The theory distinguishes between those embedded disjunctions that should yield illusions and those that should not. In Experiment 1, we corroborated this distinction. In Experiment 2, we extended the illusory inferences to a more stringently controlled set of problems. In Experiment 3, we established a novel method for reducing illusions by calling for participants to make auxiliary inferences.

Individuals with no training in logic are able to make valid deductions. Consider, for instance, the following inference:

Either memory is full or the server is busy, but not both.
Memory is not full.
Therefore, the server is busy.

Even if you know nothing about memory or servers, you can grasp that the conclusion follows from the premises; that is, the inference is *valid*: If its premises are true, its conclusion must be true too (Jeffrey, 1981). The ability to make valid deductions is a cornerstone of rationality, yet no consensus exists about the logical competence of naive individuals (i.e., those with no training in logic). Some theorists have even argued that deductive ability is not part of their thinking. In the case of inferences based on quantifiers such as *all* and *some*, individuals are instead supposed to rely on the "atmosphere" of the premises: They select conclusions that match the quantifiers in the premises (e.g., Wetherick & Gilhooly, 1995). Another view is that deductive validity is the wrong criterion to assess rationality, because everyday reasoning is probabilistic (e.g., Oaksford & Chater, 2001, p. 349). Similarly, Hertwig, Ortmann, and Gigerenzer (1997, pp. 105–106) wrote that those who study first-order logic or variants thereof, such as mental rules and mental models, ignore the ecological and social structure of environments." Still others argue that deductive reasoning depends on pragmatic schemas for specific contents (e.g., Cheng & Holyoak, 1985) or on innate modules adapted to deal with specific contents, such as checking for cheaters (e.g., Cosmides, Tooby, Fiddick, & Bryant, 2005). One counterexample to all of these views is the worldwide popularity of Sudoku puzzles. Their solutions depend solely on pure deduction, and so it is useless to rely on atmosphere, probabilities, fast and frugal heuristics, or content-specific modules. Naive individuals soon learn that Sudoku puzzles, which are quite remote from the ecological structure of daily life, call for deduction (Lee, Goodwin, & Johnson-Laird, 2008). As the popularity of these puzzles shows, individuals enjoy exercising this abstract ability, which is presumably a prerequisite for the development of logic, math, and science.

Psychological theories of deduction fall into two broad categories: those based on formal rules akin to those in the proof theory of logic (e.g., Braine & O'Brien, 1998; Rips, 1994) and those based on models akin to those in the semantic theory of logic (e.g., Johnson-Laird & Byrne, 1991; Polk & Newell, 1995). Formal rule theories postulate that deduction is a process of proof in which the rules are used to derive conclusions from the premises. A precursor to reasoning is, accordingly, the recovery of the logical form of premises to allow the application of rules. In logic, the logical form of a sentence is transparent, because it is defined by the syntax of the formal language; in daily life, sentences themselves are not normally true or false; rather, the propositions that they are used to express are. Hence, a major and unsolved problem is the recovery of the logical form of propositions. Rips's (1994) computer implementation of his theory finesses the problem by calling for users to input the logical forms of premises.

Modern logicians are often skeptical about the role of logical form in everyday reasoning. Barwise (1989, p. 4),

S. Khemlani, khemlani@princeton.edu

for example, wrote, "I find the notion [of logical form] unilluminating. Within the model-theoretic tradition, valid entailments are valid not in virtue of form, but in virtue of content." The model-theoretic tradition to which he refers relies on abstract models, such as truth tables, to specify the meanings of logical terms and to determine the validity of inferences that hinge on them. The theory of mental models, which was inspired in part by this tradition in logic (Johnson-Laird, 1983), is based on the notion that individuals reason, both deductively and inductively, on the basis of possibilities. They use the meanings of sentences and their knowledge to envisage what is possible, given the propositions expressed in the premises, and they represent these possibilities in mental models. A conclusion that holds in all of these models is *necessary*; a conclusion that holds in at least one model is *possible*; and a conclusion that holds in most models is *probable* (Johnson-Laird & Byrne, 1991). Likewise, a conclusion is invalid if it has a counterexample—that is, a possibility in which the premises hold but the conclusion does not. Mental models differ from other proposed sorts of mental representation, because models are as iconic as possible: Their structures correspond to the structure of what they represent. They can likewise unfold in time kinematically to simulate sequences of events (Johnson-Laird, 1983). But, they can also contain symbols that are not iconic, such as a symbol for negation (Khemlani, Orenes, & Johnson-Laird, 2009). The model theory provides an explanation of how individuals make deductions, inductions, explanatory abductions, probabilistic inferences, and inferences to default conclusions that hold in the absence of evidence to the contrary (Johnson-Laird, 2006). In the next section, we describe in more detail the operation of forming models, and show how it leads to the prediction of systematic fallacies that are compelling—for example, so-called *illusory* inferences. Previous studies have demonstrated their occurrence, but they have tended to rely on conditionals of the form *if A then B* (Johnson-Laird & Savary, 1999). The meanings of conditionals are controversial, and critics have argued that the illusory conclusions are, in fact, valid (Handley, Evans, & Thompson, 2006; Rips, 1997; Stenning & van Lambalgen, 2008). Although we do not accept this argument, in our present study, we used novel instances of illusory inferences that are simpler than those previously studied and that use connectives with uncontroversial meanings, such as disjunctions and conjunctions. In the article, we also examine ways to reduce the illusions.

Mental Models and Illusions

The inference with which we began was based on these premises:

Either memory is full or the server is busy, but not both.
Memory is not full.

The mental models of the first premise, which is an exclusive disjunction, represent the two sorts of possibility in which the premise holds, shown here on separate lines:

Memory full	
	Server busy

For convenience, we use these expressions to denote models, but real models have an iconic structure. The second premise eliminates the first of these models, and so the premises hold only in the second model. It follows validly that:

Therefore, the server is busy.

The conclusion is valid because there is no model of the two premises in which it is false. Theories based on formal rules also yield a proof for the conclusion. Hence, to discriminate between the two sorts of theory, we need to examine differential predictions about the same inference. One such prediction arises in the case of illusory inferences, which are a prediction unique to the model theory.

The model theory postulates a principle of *truth*: Mental models represent what is true and not what is false. This principle minimizes both the processing load on working memory and the number of resulting models. But the principle is subtle, because it applies at two levels. At the first level, mental models represent only those possibilities in which premises hold. At the second level, a mental model of a possibility represents those clauses in the premises, whether they are affirmative or negative, only when they hold. The mental models of the disjunction above illustrate both points. The two models representing the exclusive disjunction are those possibilities in which the disjunction holds, and each possibility represents the status of only one of the two clauses—the possibility in which the clause is true. *Mental* models therefore contrast with *fully explicit* models, which represent the status of all of the clauses in the premises in each model. The fully explicit models of the disjunctive premise above are as follows:

Memory full	¬Server busy
¬Memory full	Server busy

where "¬" is the symbol for negation, and so false propositions are represented by their negations. Individuals normally rely on mental models when they reason, but they can enumerate the fully explicit models of simple propositions.

A computer program (written in Common LISP) implements the principle of truth (Johnson-Laird & Savary, 1999). The program takes as input a set of premises, which may contain conjunctions, inclusive and exclusive disjunctions, and various other connectives, and its output is a set of mental models representing the possibilities consistent with the premises. The program also constructs fully explicit models, and it returns a minimal description of them. To our surprise, we discovered in the output of the program that the mental models of a set of premises do not always correspond to the fully explicit models of the premises. If reasoning is based on mental models, then individuals should make systematic fallacies in reasoning from certain sets of premises. In order to elucidate this prediction, we consider the workings of the program in more detail.

Any sentential connective in logic can be analyzed in terms of negation and conjunction; for example, a disjunction of the form *A or B or both* is equivalent to *not(not A and not B)*. Hence, in order to understand how mental models of premises are constructed, it is necessary to understand only negation and conjunction. The negation of a set of models is simply the complement of those models. Hence, the negation of the disjunction above,

It is not the case that either memory is full or the server is busy

has the following fully explicit models:

Memory full	¬Server busy
¬Memory full	Server busy

These models are indeed the complement of those above.

The conjunction of two sets of models calls for the pairwise conjunction of each model in one set with each model in the other set. These conjunctions are obvious in the case of fully explicit models. The conjunction of two models such as

A B and B C

yields the following model:

A B C

But when one model represents a proposition and the other model represents its negation, the result is the null model (akin to the empty set), which represents a contradiction; for example,

A B and ¬B C

yields

nil

Likewise, the conjunction of nil with any model also yields nil; for example,

nil and D E

yields

nil

The rules for the conjunction of mental models rely on the same three procedures, but they call for an additional procedure. What happens if two mental models to be conjoined contain no items in common? This puzzle arises in forming the conjunction of the mental models of our initial example. The models of the disjunctive premise

Memory full	
	Server busy

have each to be conjoined with the model of the second premise

¬Memory full

The first conjunction

Memory full and ¬Memory full

yields the null model. But, the conjunction

Server busy and ¬Memory full

is problematic. The model theory treats the absence of a proposition in a model as equivalent to its negation if that proposition occurs elsewhere in the same set of models. In the preceding conjunction, the proposition *Memory full* does occur in the same set off models as *Server busy*, and so its absence is treated as its negation. Hence, the conjunction becomes

¬Memory full Server busy and ¬Memory full

The result is the model

¬Memory full Server busy

and so it follows that the server is busy. The four principles for conjoining models, which we have now illustrated, are summarized in Table 1.

Table 1
The Procedures for Conjoining Pairs of Mental Models and Pairs of Fully Explicit Models

1. If one mental model represents a proposition, *A*, which is not represented in the second mental model, and *A* occurs in at least one of the sets of models from which the second model is drawn, its absence in the second model is treated as its negation (and Procedure 2 below applies); otherwise its absence is treated as its affirmation (and Procedure 3 below applies). This procedure applies only to mental models.
2. The conjunction of a pair of models containing, respectively, a proposition and its negation yields the null model (of an impossible instance); for example,
   A B and ¬A B yield nil.
3. The conjunction of a pair of models that are not contradictory yields a model representing all of the properties in the models; for example,
   A B and B C yield A B C.
4. The conjunction of a null model with any model yields the null model; for example,
   A B and nil yield nil.

The principle of truth has an unexpected consequence, which we discovered in the output of the program. It predicts that, in certain cases, individuals should make systematic invalid inferences. Consider, for instance, the following assertion:

You have the bread, or else you have the soup or else the salad.

The mental models of assertion represent three possibilities for what one has:

Bread		
	Soup	
		Salad

Given the further premise

You have the bread,

these models imply that one cannot have the soup, the salad, or both of them. But this inference is invalid, as is shown by the fully explicit models of the two premises. They show that, when it is true that one has the bread, the proposition that one has the soup or else the salad is false. And there are two ways in which this latter proposition can be false: One way is if one has neither the soup nor the salad, and the other way is if one has both the soup and the salad. Hence, when one has the bread, one can have both the soup and the salad. The fully explicit models of two premises are, accordingly,

Bread	¬Soup	¬Salad
Bread	Soup	Salad

This example illustrates an illusory inference. Most individuals should infer that, given that they have the soup, they cannot have the salad and the bread. The inference seems compelling, but it is invalid.

Illusory inferences put formal rule theories on the horns of dilemma. On one hand, theories based on rules that yield only valid inferences cannot explain the illusions. On the other hand, if rules are introduced to account for the illusions, the system is at risk of inconsistency, because it contains both correct and incorrect rules. Likewise, theories of reasoning based on probabilistic methods are unable to predict illusory inferences. These theories postulate that individuals infer conclusions because they are probable. But, if individuals infer that a conclusion is impossible when, in fact, it has a nonzero probability of occurrence, the theories can offer no account of their performance. As we show, individuals do indeed make such errors in a systematic way. In summary, illusory inferences appear to be a decisive test for the use of mental models, because no other current theory predicts their occurrence.

EXPERIMENT 1
Disjunctive Illusions

In Experiment 1, we examined four sorts of inference: two experimental problems that should yield illusions and two control problems that should not. All four problems were based on disjunctions, and the first illusory problem (exclusive–exclusive) was as follows.

> Suppose that only one of the following assertions is true:
>
> (1) You have the mints.
> (2) You have the gumballs or the lollipops, but not both.
>
> Also, suppose you have the mints. What, if anything, follows? Is it possible that you also have either the gumballs or the lollipops? Could you have both?

This problem is of the same sort as our earlier example. The rubric “only one of the following assertions is true” establishes an exclusive disjunction between Assertions 1 and 2, and Assertion 2 is in turn an exclusive disjunction. The premises therefore yield three mental models of the candies that you can have:

Mints		
	Gumballs	
		Lollipops

Given the further premise *you have the mints*, the participants should respond “no” to the question about whether they can have both of the other candies. The mental models, however, fail to represent that, when one assertion is true, the other assertion is false. The fully explicit models show that the preceding response is an illusion:

Mints	¬Gumballs	¬Lollipops
Mints	Gumballs	Lollipops

Given the premise *you have the mints*, one way in which Assertion 2 can be false is if one also has both the gumballs and the lollipops.

The first control problem (inclusive–inclusive) was as follows.

> Suppose that at least one of the following assertions is true, and possibly both:
>
> (1) You have the marshmallows.
> (2) You have the truffles or the Jolly Ranchers, and possibly both.
>
> Also, suppose you have the marshmallows. What, if anything, follows? Is it possible that you also have either the truffles or Jolly Ranchers? Could you have both?

The mental models of the problem allow one to have any candy by itself, any pair of candies, or all three of them. Hence, reasoners should respond that it is possible to have all three candies. The fully explicit models support the same conclusion. Table 2 summarizes these two problems and the other two sorts of problems. As the table shows, two problems should give rise to illusory inferences, and two should give rise to correct inferences.

Method

Participants. Eighteen undergraduates from Princeton University served as their own controls and carried out all four problems. None of the participants had received any training in logic. Each participant received the problems in a different random order.

Procedure and Materials. The experiment was carried out over the Internet. The participants were told that they would be asked to infer conclusions about what types of candy they could have from a candy shop. They were asked to type out their conclusions online, and they were told that they could take as long as they needed to complete the task. They were asked not to use pen and paper or any external method to solve the problems. Each problem referred to a different set of candies.

Results

Table 2 presents the percentages of participants who responded correctly to the question *Could you have both candies?* As the table shows, the participants performed much more accurately with the control problems than with the illusions: 16 out of the 18 participants showed this difference (binomial test, $p < .0001$). This result corroborates the model theory’s predictions. The difference between the two sorts of problems cannot be explained in terms of the numbers of mental models or the number of fully explicit models in which the assertions hold (see Table 2). Yet, the premises differed over the four sorts of

Table 2
The Four Sorts of Problems in Experiment 1, Their Mental Models, Their Fully Explicit Models, and the Percentages of Participants Who Responded Correctly to the Question About the Conjunction

Type of Problem	Description	Mental Models			Fully Explicit Models			Correct Answer	Percentage Correct
Control	Inclusive–inclusive	A			A	¬B	¬C	Yes	100
	(A or (B or C))		B		¬A	B	¬C		
				C	¬A	¬B	C		
			B	C	¬A	B	C		
		A	B		A	B	¬C		
		A		C	A	¬B	C		
		A	B	C	A	B	C		
Control	Exclusive–inclusive	A			A	¬B	¬C	Yes	78
	(A or else (B or C))		B		¬A	B	¬C		
				C	¬A	¬B	C		
			B	C	¬A	B	C		
Illusory	Inclusive–exclusive	A			A	¬B	¬C	Yes	22
	(A or (B or else C))		B		A	B	C		
				C	¬A	B	¬C		
		A	B		¬A	¬B	C		
		A		C	A	B	¬C		
					A	¬B	C		
Illusory	Exclusive–exclusive	A			A	¬B	¬C	Yes	17
	(A or else (B or else C))		B		A	B	C		
				C	¬A	B	¬C		
					¬A	¬B	C		

Note—The question about the conjunction was the following: *Suppose A. What, if anything, follows? Is it possible that B or C is true? Is it possible that both B and C are true?*

problems, and so some unknown factor could have been responsible for the difference in accuracy.

EXPERIMENT 2
A Corroboration and the Effect of Instructions

Unlike in Experiment 1, in Experiment 2, we examined illusory and control problems based on the same premises. Each problem was presented with two separate questions on separate trials with different contents: One question was predicted to elicit an illusory inference, and one question was a control predicted to elicit a correct answer. Consider these premises based on two exclusive disjunctions:

> Suppose that one of the following assertions is true and one is false.
>
> (1) You have the blue candies and the red candies.
> (2) You have the red candies or else the orange candies but not both.

The rubric makes very clear that there is an exclusive disjunction between Assertions 1 and 2. Hence, the premises yield the following mental models of the candies:

Blue	Red	
	Red	
		Orange

The following question should therefore yield an illusory inference:

> Is it possible to have only the blue candies and the orange candies?

The mental models yield the response "no." But, for the control question

> Is it possible to have only red candies?

the participants should respond "yes." The fully explicit models show that the first response is illusory, because one way in which the conjunction in the first premise can be false is if one has blue candies but not red ones, and one way in which the disjunction in the second premise can be true is if one has orange candies but not red ones. In the experiment, we also tested whether special instructions might remediate the illusions. Halfway through the experiment, one group of participants was given instructions designed to get them to think about both what was true and what was false (see Yang & Johnson-Laird, 2000).

Method

Participants and Design. Seventeen students at Princeton University were tested individually, and each participant completed two blocks of four problems. Table 3 presents the four sorts of problems, which were each presented with a control question and in a separate block with an illusory question. Each participant carried out two blocks of trials with the four sorts of problems. In each block, two of the problems occurred with a control question and two of them occurred with an illusory question. We tested two separate groups of participants. One group was given instructions designed to reduce the illusions after they had carried out the first block of problems (the *instructed* group). The other group was not given these instructions (the *uninstructed* group).

Procedure and Materials. Each problem referred to three different sorts of candy, and no participant encountered the same set of candies more than once. A preliminary version of the experiment was carried out over the Internet but did not yield the remedial effect of

Table 3
The Four Sorts of Problems in Experiment 2, Their Mental Models, Their Fully Explicit Models, and the Percentages of Correct Answers to the Control and Illusory Questions

Problem	Mental Models	Fully Explicit Models	Question	Correct Answer	Percentage Correct	
					Instructed	Uninstructed
Exclusive–exclusive *One is true, and one is false:* *(1) A and B, (2) B or else C.*	A B B C	A B C ¬A B ¬C A ¬B C ¬A ¬B C	Control: *Is it possible to have only B?*	Yes	90	100
			Illusion: *Is it possible to have only A and B?*	No	40	0
Exclusive–inclusive *One is true, and one is false:* *(1) A and B, (2) B or C.*	A B B C B C	¬A B ¬C A ¬B C ¬A ¬B C ¬A B C	Control: *Is it possible to have only B?*	Yes	100	93
			Illusion: *Is it possible to have A and B?*	No	20	7
Inclusive–exclusive *One or both are true:* *(1) A and B, (2) B or else C.*	B C A B A B C	A B C ¬A B ¬C A ¬B C ¬A ¬B C A B ¬C	Control: *Is it possible to have only B and C?*	No	80	93
			Illusion: *Is it possible to have only A and C?*	Yes	20	21
Inclusive–inclusive *One or both are true:* *(1) A and B, (2) B or C.*	B C B C A B A B C	¬A B ¬C A ¬B C ¬A ¬B C ¬A B C A B ¬C A B C	Control: *Is it possible to not have A, B, or C?*	No	100	100
			Illusion: *Is it possible to have only A and C?*	Yes	30	21

instructions, probably because the participants failed to follow them in detail. The present experiment was accordingly carried out in the laboratory. The participants were told that they would have to infer conclusions about whether or not it was possible to have certain sorts of candy. They responded by selecting a button for *yes* or for *no*. They were told that they could take as long as they needed to complete the task. They also rated how confident they were in their responses on a 7-point scale (1 meant *not confident at all*, and 7 meant *very confident*).

The key instructions for the instructed group were:

> To solve these problems correctly, you need to do the following:
>
> (1) Mentally select your response.
>
> (2) Go back and check whether your response preserves the relationship between the premises. For instance, if you are told that one of the premises is true and one is false, you need to make sure that your response takes both these facts into account.

The instructions then presented a worked example in which one premise was true and one premise was false but for a problem that did not occur in the experiment. The uninstructed group took an unrelated survey between the two blocks of trials, which took about the same time as did the remedial instructions.

Results

Table 3 presents the percentages of correct responses for the two groups. The participants made a much greater percentage of correct responses to the control questions (95%) than to the illusory questions (19%, Wilcoxon test, $z = 8.42$, $p < .000001$). Table 4 presents the results for the two blocks for each group. As it shows, mean performance appeared to increase from the first block to the second block, but the tendency was unreliable (Wilcoxon test, $z = .40$, $p = .69$). The remedial instructions occurred after the first block, and the results supported their efficacy. In the second block, the difference between the two sorts of problem (the controls and the illusions) was smaller for the instructed group than for the uninstructed group (34% vs. 79%; Mann–Whitney test, $z = 1.71$, $p < .05$). Indeed, in the first block of trials, there was no reliable difference in accuracy on the illusions between the two groups (Mann–Whitney test, $z = .33$, $p = .74$), but the difference in the second block was reliable (Mann–Whitney test, $z = 2.44$, $p < .01$). The participants' mean confidence in their responses was high: Their mean ratings ranged on the 7-point scale from 5.45 to 6.26, and there was no reliable difference among them. Hence, the instructions yielded an improvement in performance on illusions.

Table 4
Percentages of Correct Responses Over the Two Blocks of Trials for the Instructed Group and the Uninstructed Group in Experiment 2

	Instructed Group		Uninstructed Group	
	Control	Illusion	Control	Illusion
First block of trials	96	13	92	16
Second block of trials	88	38	100	8

EXPERIMENT 3
The Modulation of Illusions With Inferences

Experiment 2 showed that instructions to consider both what is true and what is false improved performance with the illusory inferences. In Experiment 3, we examined two manipulations that should also affect performance with the illusions. The first manipulation compared propositions referring to properties (or *unary* relations), such as *one course is meat*, with propositions referring to binary relations, such as *Lisa is in front of Maggie*. The two sorts of proposition accordingly differ in relational complexity (Halford, 1993), which is known to affect the difficulty of reasoning (Birney, Halford, & Andrews, 2006; Goodwin & Johnson-Laird, 2005).

The second manipulation was whether or not the participants had to evaluate a conclusion requiring an extra inference from the premises. This inference called for the participants to think both about what is true and about what is false, and so it should also make the participants less susceptible to the illusions. In the case of premises about properties, a preliminary assertion provided a definition, and the inference to be evaluated depended on this definition. For example, the following problem concerned a meal in a restaurant:

> An entrée consists of one meat and one vegetable course. Suppose that one of the following assertions is true and one is false.
>
> (1) One course is meat, and one course is salad.
> (2) One course is salad, or else one course is vegetables.
>
> Is an entrée possible?

The mental models of the premises yield the following possibilities for a meal:

Meat	Salad	
	Salad	
		Vegetable

These models imply that the answer to the question is, “No, an entrée is not possible.” But, as before, if Assertion 1 is true, one way in which Assertion 2 can be false is that both of its propositions are true, and so *meat*, *salad*, and *vegetable* are possible, and the correct answer is, “Yes, the entrée is possible.” Indeed, the fully explicit models of the possible meals are as follows:

Meat	Salad	Vegetable
¬Meat	Salad	¬Vegetable
Meat	¬Salad	Vegetable
¬Meat	¬Salad	Vegetable

The experiment contrasted these problems with cases—akin to those in the previous experiments—in which the conclusion to be evaluated merely listed a conjunction of propositions (e.g., “Is it possible that one course is meat, one course is salad, and one course is vegetables?”). It may seem counterintuitive that an additional inference might improve performance, but the definition of *entrée* does call for the participants to explicitly consider both assertions, and so they should be more likely to consider the truth of one and the falsity of the other.

For the problems about relations, the question posing an additional inference also called for the participants to consider what is true and what is false; for example,

> Suppose that one of the following assertions is true and one is false:
>
> (1) Lisa is in front of Maggie, and Maggie is in front of Bart.
> (2) Maggie is in front of Bart, or else Bart is in front of Homer.
>
> Is it possible that Lisa is in front of Homer?

As in the previous example, the mental models yield the answer “No,” but the fully explicit models show that the answer is wrong. For example, if Assertion 1 is true, then one way in which Assertion 2 could be false is that both of its propositions are true, and so, given the truth of all four propositions, Lisa is in front of Homer. The condition in which the extra inference did not occur called for the evaluation of a conjunction of clauses in the premises (e.g., “Is it possible that Lisa is in front of Maggie, Maggie is in front of Bart, and Bart is in front of Homer?”).

Method

Participants and Design. Forty-one members of the Princeton University community served as their own controls and carried out eight problems. In the experiment, we manipulated the type of assertions (property vs. relation) and whether or not an extra inference was required in order to answer the question. The extra inference was required for problems referring to a newly defined term and for problems with a transitive relation between the first and the last terms in their premises. We used two sorts of illusory problem, an inclusive–exclusive one and an exclusive–exclusive one, both of which were predicted to yield erroneous responses of “no.” Hence, the experiment also included four control problems from the previous study, which had reliably yielded the correct responses of “yes.” The problems were given in a different random order to each participant.

Procedure. The experiment was carried out over the Internet, and the participants were told that they would have to infer conclusions about conflicting situations. They responded by typing their answers into a text box on the screen. They were told that they could take as long as they needed to complete the task. They also rated how confident they were in their responses on a 7-point scale.

Results

The participants were more accurate in responding to the control problems (81% correct) than to the illusory problems (63% correct; Wilcoxon test, $z = 3.03$, $p < .005$). The illusions had a much smaller effect than did those in the previous experiments, presumably because the participants had only to answer “yes” or “no” to questions about a correct conclusion, and so the chance probability of a correct answer was 50%. There was no reliable difference between the exclusive–exclusive and inclusive–exclusive illusory problems, and so we have pooled their data in Table 5, which presents the percentages of correct responses to the illusory problems of the four sorts. Overall, there was no reliable effect of whether the participants had to make an extra inference (65% correct vs. 61% correct, Wilcoxon test, $z = .85$, $p = .39$). Problems based on properties were easier than those based on relations (69% correct vs. 57% correct, Wilcoxon test, $z = 2.83$, $p < .005$). But, as Table 5 shows, this effect merely reflected an interaction, which was highly reliable: The extra inference improved performance with properties,

Table 5
Percentages of Correct Responses to the Illusory Inferences in Experiment 3, Depending on Whether the Problems Concerned Properties or Binary Relations, and on Whether Responses Called for an Additional Inference

	No Extra Inference	Extra Inference
Properties	61	78
Binary relations	61	52

but impaired it with relations (Wilcoxon test, $z = 3.45$, $p < .001$). The extra inference reliably improved reasoning about properties (Wilcoxon test, $z = 2.33$, $p < .01$), but the deterioration in reasoning about relations was not reliable (Wilcoxon test, $z = 1.22$, $p = .22$).

The interaction was unexpected, but the following post hoc explanation seems plausible. For the assertions about properties, the extra inference depends on the same number of possibilities as does the condition with no extra inference. So, for example, the exclusive disjunction of

(1) One course is meat, and one course is salad.
(2) One course is salad, or else one course is vegetables.

yields these possibilities

Meat	Salad	Vegetables	(1) is true, (2) is false.
	Salad		(1) is false, (2) is true.
Meat		Vegetables	(1) is false, (2) is true.

For the extra inference, individuals ought to conclude that an *entrée* (*meat* and *vegetables*) is possible; and for the other condition, they can conclude that *meat*, *salad*, and *vegetables* are possible. But the extra inference contains a conclusion referring to items in both premises, and so the participants should be more likely to consider the truth of one premise and the falsity of the other. In contrast, for the assertions based on relations, there are many more possibilities to take into account in assessing the extra inference. For example, the exclusive disjunction of the following:

(1) Lisa is in front of Maggie and Maggie is in front of Bart.
(2) Maggie is in front of Bart or else Bart is in front of Homer.

The inferential question is whether it is possible that Lisa is in front of Homer, and this relation holds in only 7 of the 14 different possible orders for the four individuals. When there was no extra inference, the question merely concerned a conjunction of the clauses, and it is not necessary to work out the possible order of the individuals in order to do so. It suffices to consider the three possibilities in which the premises hold, and to infer that it is possible that Lisa is in front of Maggie, Maggie is in front of Bart, and Bart is in front of Homer. In short, for properties, the number of possibilities to consider remains the same whether or not there is an extra inference, and so the extra inference helps participants to consider the truth of one premise and the falsity of the other. But, for relations, this aspect of the extra inference is offset by the much greater number of possibilities that participants need to consider.

GENERAL DISCUSSION

The present study shows that robust illusory inferences occur with premises in which one disjunction is embedded in another. The meanings of exclusive and inclusive disjunctions are easy for individuals to grasp when only a single disjunction occurs in a premise. The theory's principle of truth, however, postulates that mental models represent what is true and not what is false. A computer program implementing this principle showed that certain disjunctive problems should yield systematically invalid conclusions. Experiment 1 showed that individuals erred with problems of this sort: *Only one of these assertions is true: (1) A, (2) B or else C*. When they were told that *A* was the case, they tended to infer that it was impossible for *B and C* to occur. The theory predicts this illusion because mental models fail to represent what is false: One way in which *B or else C* can be false is if both *B* and *C* occur. In Experiment 2, we extended these results by showing that illusions remain reliably harder than control problems when both sorts of problem are based on the same premises and only the questions differ. For example, premises of the sort *One of these assertions is true and one is false: (1) A and B, (2) B or else C* yield these mental models:

A	B	
	B	
		C

And most participants accordingly inferred wrongly that *A and B* alone was possible, but rightly that *B* alone was possible. The fully explicit models (see Table 3) show that *A and B* can occur only with *C*. The occurrence of the conjunction in the illusory inference cannot account for the error, because some control problems, to which the participants responded correctly, also had questions based on conjunctions (see Table 3).

In a preliminary experiment carried out online, we examined whether the instruction to think carefully about both what is true and what is false ameliorated the illusions. It failed to do so, despite an effect with illusions based on quantifiers such as *all* and *some* (Yang & Johnson-Laird, 2000). We surmised that the participants had not followed the instructions properly, and therefore, Experiment 2 was a face-to-face replication. The instructions yielded the predicted improvement in performance with the illusory inferences.

In Experiment 3, we examined two other potential ways to affect the illusions. One way manipulated the relational complexity of the premises (Halford, Wilson, & Phillips, 1998), contrasting those about properties, such as *the course is soup*, with those about relations, such as *Lisa is in front of Maggie*. The other way contrasted conclusions that called for an extra inference with those that did not. The principal result was an unexpected interaction. The extra inference reliably improved performance with inferences about properties but tended to impair inferences about binary relations. An extra inference ought to impair reasoning (cf. Johnson-Laird & Byrne, 1991; Rips, 1994), but since it forced individuals to consider the joint implications of the truth of one premise and the falsity of the other premise, it improved performance. This advantage, however, was offset in inferences about relations, because the extra inference called for them to consider many more possibilities than they did in the condition in which it was not required.

In conclusion, the occurrence of illusory inferences is a litmus test for the use of mental models. It is particularly telling that the sets of mental models in Experiment 2 predicted that some questions would be easy to answer correctly, whereas other questions would elicit systematically invalid inferences. No other current theory of reasoning, whether based on formal rules of inference (e.g., Braine & O'Brien, 1998; Rips, 1994) or on the probability calculus (Oaksford & Chater, 2001), predicts the illusions or the remedial effects of instructions to think about truth and falsity. These theories could introduce fallacious rules in order to predict the illusions, but the resulting system is likely to be inconsistent and, in consequence, to predict other sorts of fallacy that do not occur. Illusory inferences can be elicited using conditionals and quantifiers, but the experiments presented here demonstrate that illusions occur in simpler cases too. The inferences in all three experiments were based on *or*, in its inclusive or exclusive sense, and sometimes concerned only three entities in the domain of discourse. Participants nevertheless succumbed to these minimal illusions.

AUTHOR NOTE

This research was supported by a National Science Foundation Graduate Research Fellowship to the first author and by National Science Foundation Grant DRMS 0844851. We are grateful to Sam Glucksberg, Adele Goldberg, Geoffrey Goodwin, Nuria Carriedo, Jeremy Boyd, Paula Rubio, Rina Ayob, and Olivia Kang for their helpful suggestions. Special thanks go to Micah Clark for his critique of the model theory, as well as to Klaus Oberauer, Juan A. García Madruga, and two anonymous reviewers for their comments on previous versions of the article. Correspondence concerning this article should be addressed to S. Khemlani, Department of Psychology, Princeton University, Princeton, NJ 08540 (e-mail: khemlani@princeton.edu).

REFERENCES

Barwise, J. (1989). *The situation in logic*. Stanford, CA: Center for the Study of Language and Information.

Birney, D. P., Halford, G. S., & Andrews, G. (2006). Measuring the influence of complexity on relational reasoning: The development of the Latin Square task. *Educational & Psychological Measurement*, **66**, 146-171.

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

Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology*, **17**, 391-416.

Cosmides, L., Tooby, J., Fiddick, L., & Bryant, G. A. (2005). Detecting cheaters. *Trends in Cognitive Sciences*, **9**, 505-506.

Goodwin, G. P., & Johnson-Laird, P. N. (2005). Reasoning about relations. *Psychological Review*, **112**, 468-493.

Halford, G. S. (1993). *Children's understanding: The development of mental models*. Hillsdale, NJ: Erlbaum.

Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral & Brain Sciences*, **21**, 803-831.

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

Hertwig, R., Ortmann, A., & Gigerenzer, G. (1997). Deductive competence: A desert devoid of content and context. *Current Psychology of Cognition*, **16**, 102-107.

Jeffrey, R. C. (1981). *Formal logic: Its scope and limits* (2nd ed.). New York: McGraw-Hill.

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

Johnson-Laird, P. N. (2006). *How we reason*. Oxford: Oxford University Press.

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

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

Khemlani, S., Orenes, I., & Johnson-Laird, P. N. (2009). *Negation: A theory of its meaning, use, and interpretation*. Manuscript submitted for publication.

Lee, N. Y. L., Goodwin, G. P., & Johnson-Laird, P. N. (2008). The psychological problem of Sudoku. *Thinking & Reasoning*, **14**, 342-364.

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

Polk, T. A., & Newell, A. (1995). Deduction as verbal reasoning. *Psychological Review*, **102**, 533-566.

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

Rips, L. J. (1997). Goals for a theory of deduction: Reply to Johnson-Laird. *Minds & Machines*, **7**, 409-424.

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

Wetherick, N. E., & Gilhooly, K. J. (1995). "Atmosphere," matching, and logic in syllogistic reasoning. *Current Psychology*, **14**, 169-178.

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

(Manuscript received June 22, 2008;
revision accepted for publication January 18, 2009.)