Illusory Inferences about Embedded Disjunctions

Sangeet Khemlani (khemlani@princeton.edu)
P.N. Johnson-Laird (phil@princeton.edu)
Department of Psychology
Princeton University
Princeton, NJ 08540 USA

Abstract

The mental model theory of reasoning postulates that individuals construct mental models of the possibilities consistent with premises, 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 amongst current theories of reasoning, because no alternative theory, whether based on formal rules of inference or on probabilistic considerations, can explain these illusory inferences. We report two studies of novel illusory inferences that depend on embedded disjunctions, e.g., premises of the form *A or else (B or else C)*. The theory distinguishes between those embedded disjunctions that should yield illusions and those that should not. Experiment 1 corroborated this distinction. Experiment 2 extended the phenomena to a more controlled set of problems.

Keywords: deductive reasoning; mental models; illusory inferences.

Introduction

Humans reason about concepts and situations, and they can do so without requiring training in logic. They are unable to introspect on the mental processes underlying their reasoning; and so the task of describing these processes is left to cognitive scientists. Much of their empirical work on the psychology of reasoning has focused on the inferences humans systematically draw and those they fail to draw, and many competing theories of reasoning have been constructed from these data (for a review, see Evans, Newstead, and Byrne, 1993).

One of these theories is that humans are fundamentally rational and reason by using formal rules of inference akin to those in symbolic logic (Rips, 1994; Braine and O'Brien, 1998). Another theory is that logic is an inappropriate normative model, and that human reasoning – even deductive reasoning – is probabilistic and best described by the probability calculus (Oaksford and Chater, 1998). Our own theory is that human reasoning, both deductive and inductive, is based on the representation of possibilities. Individuals construct *mental models* of the possibilities consistent with the premises. In deduction, a conclusion that holds in all these models is necessarily the case, a conclusion that holds in at least one model is possible, and a conclusion that holds in most models is probable (Johnson-Laird, 2006; Johnson-Laird and Byrne, 1991).

Much controversy exists about the merits of these alternative theories, but what may be most revealing about them are the situations in which human reasoners make systematically invalid inferences. We examine such erroneous patterns of reasoning in the present article.

A Model Theoretic Account of Illusory Inferences

A fundamental assumption of the model theory is the *principle of truth*: in order not to overload working memory, mental models represent only what is true and not what is false. The principle applies both to sentences as a whole, but also to clauses within them. Hence, an exclusive disjunction: *A or else not-B*, where we use 'or else' to symbolize exclusivity, has the following two mental models, in which each line represents a possibility, and '¬' represents negation:

A
¬ B

An inclusive disjunction, which we symbolize as: *A or not-B*, allows the additional possibility in which both clauses are true: A ¬ B. In certain circumstances in which a task is not too burdensome, individuals can flesh out mental models into fully explicit models, which represent clauses when they are true and when they are false. Those for the exclusive disjunction above are as follows:

A B
¬ A ¬ B

where a negation is used to represent a false affirmative, and an affirmative is used to represent a false negation.

The principle of truth seems innocuous, but it yields the unexpected prediction that individuals should make invalid inferences in certain cases. Previous studies have shown that such illusions do occur (e.g., Johnson-Laird & Savary, 1999). The present studies established a novel sort of illusion that occurs with premises based solely on disjunctions. They therefore counter the criticism that illusions arise from participants mistakenly interpreting conditionals (Rips, 1997). The next section describes some of these putative illusions, and the paper then describes our experimental investigations of them.

Illusions with Embedded Disjunctions

The inferences that we investigated are based on premises in which one disjunction is embedded within another. Our first experiment examined four sorts of problem. A computer program implementing the principle of truth predicted that two of them should yield illusory inferences, and that two of them should yield valid inferences.

The first sort of illusory problem was as follows, where the rubric 'suppose only one of the following assertions is true' has the force of an exclusive disjunction:

Illusory problem 1: Exclusive-exclusive disjunctions
Suppose 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?

The mental models of the problem allow for three possibilities, which we lay out on separate rows:

mints
gumballs
lollipops

Granted that you have the mints, it follows – as the first of these models shows – that you can have neither the gumballs nor the lollipops. The mental models, however, fail to represent that when one assertion is true the other assertion is false. Granted that you have the mints, then assertion 2 is false. And one way in which this proposition can be false is that you have neither the gumballs nor the lollipops, but another way in which it can false is that you have both the gumballs and the lollipops. So, the fully explicit models of the two possibilities consistent with the problem are as follows:

mints ¬ gumballs ¬ lollipops
mints gumballs lollipops

Hence, it *is* possible that you have all three: the mints, the gumballs, and the lollipops. But, naïve individuals should rely on mental models of the problem rather than fully explicit models and so they should err.

The second sort of illusory problem is as follows, where the rubric is equivalent to an inclusive disjunction:

Illusory problem 2: Inclusive-exclusive disjunctions
Suppose at least one of the following assertions is true, and possibly both:
1. You have the jellybeans
2. You have the peppermints or the gummy bears, but not both

Also, suppose you have the jellybeans. What, if anything, follows?

Given that you have the jellybeans, the mental models of the problem allow that you can also have either the peppermints or else the gummy bears, but not both. But, the fully explicit models allow that you have all three candies for the same reason as before: assertion 2 can be false because you have both the peppermints and the gummy bears.

The third sort of problem is a control, because its mental models yield the correct conclusion:

Control problem 1: Inclusive-Inclusive disjunctions
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?

The mental models of the problem allow you 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.

The fourth and final problem is also a control:

Control problem 2: Exclusive-Inclusive disjunctions
Suppose only one of the following assertions is true:
1. You have the gumdrops
2. You have the tootsie rolls or the gobstoppers, and possibly both

Also, suppose you have the gumdrops. What, if anything, follows?

Given that you have the gumdrops, the mental models of the problem do not allow you to have either of the other two candies. The fully explicit models likewise yield this response, and so naïve individuals should make the correct response. We used all four problems in the following experiment.

Experiment 1

Method

Design. 18 undergraduates from Princeton University served as participants, and they each carried out the four problems described above, two of which should yield illusions and two of which should yield correct responses. None of the participants had received any training in logic. The four problems were presented in a different random order to each participant.

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 in a form provided on line, and they were told that they could take as long as they needed to complete the task. They were asked not to use paper or any external method to solve the problems.

The experiment used questions that were more explicit than those described in the previous section:

What, if anything, follows? Is it possible that you also have either Candy Y or Candy Z? Could you have both?

The last question in the prompt above makes clear that participants might be able to have both candies. Each of the four problems referred to different sets of candies.

Results and discussion

Table 1 presents the percentages of participants who responded correctly to the question: *Could you have both (candy B and candy C)?* A negative response is correct only for the exclusive-inclusive control problem. But, the participants tended to succumb to the illusions and to make this response to the two illusory problems. 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, $p < .005$).

Table 1: The percentages of correct responses in Experiment 1 to the question: *Could you have both B and C?*

Control problems	%	Illusory problems	%
Inclusive-inclusive *(A or (B or C))*	100	Exclusive-exclusive *(A or else (B or else C))*	17
Exclusive-inclusive *(A or else (B or C))*	78	Inclusive-exclusive *(A or (B or else C))*	22

The illusory problems produced responses that are compelling, almost universal, but fallacious. They accordingly corroborate the predictions of the model theory. Could the results be explained in terms of the difficulty of understanding disjunctions? In straightforward deductions from a premise containing a single disjunction, exclusive disjunctions are easier than inclusive disjunctions (Bauer & Johnson-Laird, 1993). The model theory predicts this difference, because exclusive disjunctions are consistent with only two possibilities (see the Introduction), whereas inclusive disjunctions are consistent with a third possibility in which both their clauses hold. (Current formal rule theories make the opposite prediction because they postulate an additional step to convert an exclusive disjunction, *A or else B*, into *A or B, & not both A & B*, see Braine & O'Brien, 1998; Rips, 1994.) The problems in the present experiment stated that *A* was true, and as a result they yielded the following numbers of mental models:

Inclusive-inclusive: 4
Exclusive-inclusive: 4
Exclusive-exclusive: 3
Inclusive-exclusive: 3

Moreover, the presentation of explicit questions may well have led the participants to work backwards from the questioned proposition to the premises, and in this way they could avoid having to construct all the models of the premises (García-Madruga et al., 2001). Hence, what causes the difficulty of the illusory problems is not the number of mental models, but the fact that they are erroneous models.

Experiment 2

The aim of the experiment was to extend our corroboration of illusory inferences to a wider set of embedded disjunctions. These new problems contained a disjunction of a conjunction and an embedded disjunction. In Experiment 1, the illusory and control problems differed, but in the present study both sorts of problem were based on the same premises. What differed was the nature of the question that was paired with the premises. Each problem occurred 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. For instance, here is a problem with a control question:

Exclusive-exclusive. Suppose 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.

Is it possible to have only red candies?

The mental models of the premises are as follows:

blue red
red
orange

Hence, the participants should respond, 'yes'. The same form of premises was also paired with an illusory question:

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

The mental models above contain no such possibility, and so the theory predicts that individuals should respond, 'no'. The response is invalid, because one way in which the conjunction can be false is if you have blue candies but not red ones, and one way in which the embedded disjunction can be true is if you have orange candies but not red ones. The experiment also tested whether a remedial procedure used in a previous study might remediate the illusions (see Yang & Johnson-Laird, 2000). Half way through the experiment, one group of participants was given instructions designed to get them to think about what was true *and* what was false.

Method

Design. 40 members of the Princeton University community served as participants, and each participant completed two blocks of four problems. The experiment used four sorts of problems (see Appendix A), which were each presented with a control question and on a separate trial with an illusory question. Each participant carried out two blocks of trials with the four sorts of problem, e.g., there was only one exclusive-exclusive problem in each block. 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 required choosing between three different sorts of candy, and no participant encountered the same set of candies more than once.

The experiment was carried out over the Internet. The participants were told that they would have to infer conclusions about whether or not it was possible to have certain sorts of candy from a candy shop in the light of stated information. 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 seven-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 into account both these facts.

The instructions then presented a worked example of a case 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 merely took a short rest of a comparable duration between the two blocks of trials.

Results and discussion

Table 2 presents the percentages of correct responses for the two groups, and Appendix A presents their breakdown by problem. The participants made more correct responses to the control questions (93%) than for the illusory questions (37%, Wilcoxon test, $z = 8.94$, $p < .0001$). There was no reliable difference between the two groups (Mann-Whitney test, $z = 1.7$, $p = 0.09$). But, the predicted interaction was reliable: the difference in accuracy between the control and illusory questions was reliably greater for the uninstructed group than for the instructed group (Mann-Whitney test, $z = 2.2$, $p < .05$).

Table 2: The percentages of correct responses over the four sorts of premises

	Instructed group		Uninstructed group	
Type of question	**Control**	**Illusion**	**Control**	**Illusion**
Exclusive-exclusive	94	38	92	17
Exclusive-inclusive	94	38	96	21
Inclusive-exclusive	88	63	92	21
Inclusive-inclusive	94	56	92	58

Table 3 shows results for the two blocks for each group. As the Table shows, a trend occurred in which performance declined from the first block to the second block, but the trend was unreliable (Wilcoxon test, $z = 1.03$, $p < .29$). In the first block of trials, there was no reliable difference in accuracy on the illusions between the two groups (Mann Whitney, $z = 1.12$, $p > 0.2$), whereas the difference in the second block was reliable (Mann Whitney test, $z = 2.38$, $p < 0.02$).

Table 3: The percentages of correct responses over the two blocks of trials

	Instructed group		Uninstructed group	
	Control	**Illusion**	**Control**	**Illusion**
1st block of trials	97	50	94	38
2nd block of trials	88	47	92	21

However, the instructions certainly did not yield a complete improvement in performance. One interpretation of the results is that the instructions helped to prevent a decline in performance, which appeared to occur in the uninstructed group. The participants' mean confidence in their responses was high: their ratings ranged on the seven-point scale from 5.08 to 6.41, and there were no reliable differences amongst them.

The principal result was that a robust difference in performance occurred between the control problems, which were easy, and the illusory problems, which were difficult. Only in the case of the inclusive-inclusive disjunction *((A and B) or (B or C))* did participants make more correct responses than incorrect responses to an illusory inference.

The method of using additional instructions to attenuate errors in reasoning has been partially successful in previous studies, though it had the unintended effect of reducing performance on control problems (Yang and Johnson-Laird, 2000). To some degree, the same phenomenon occurred in the present study, except that the instructions did not yield an improvement in performance with the illusory questions.

General Discussion

Our results show that robust illusory inferences occur with premises in which one disjunction is embedded within another. The meaning of exclusive and inclusive disjunctions is easy for individuals to grasp when only a single disjunction occurs in a premise. The model theory, however, predicts that individuals tend to represent only what is true at the expense of what is false. When a computer program applied this principle to problems in which one disjunction is embedded within another, the program predicted that certain problems should yield systematically invalid conclusions. Experiment 1 showed that individuals erred with problems of the form: *A or else (B or else C)* when they were told that *A* was the case. They tended to infer that it was impossible for *A, B, and C*, all to hold, and the theory predicts this illusion because mental models fail to represent what is false and, in particular, that one way in which *B or else C* can be false is if both *B and C* are true. This systematic fallacy cannot be predicted by theories based on formal rules of inference (Braine & O'Brien, 1998; Rips, 1994), because these theories rely on valid rules, and so they have no way to account for systematic fallacies. Likewise, theories that assume that the probability calculus describes our inferences cannot explain the error, either, because *B and C* has a definite probability

of occurring given the truth of the premises (cf. Oaksford & Chater, 1998).

Experiment 1 used different sorts of problem for the illusions and the controls, and so it is conceivable that some other aspect of the problems was responsible for the pattern of results. Experiment 2, however, used the same premises for both the illusions and the control problems. They differed solely in the questions that occurred with them, which either concerned single propositions or conjunctions.

Experiment 2 also examined the remedial effect of instructions to consider both what is true and what is false. These instructions did not improve performance overall with the illusory problems, but they may have offset a trend for performance to be poorer in the second block of problems than in the first block of problems. Instructions of this sort have worked in at least one previous study (Yang & Johnson-Laird, 2000). One reason for their lack of efficacy in Experiment 2 may have been the reluctance of the participants to put the instructions into effect, especially as the experiment was carried out online.

Illusory inferences can be elicited using conditionals and quantifiers, but the studies presented here demonstrate that illusions occur in simpler cases too. The problems in both experiments were based on 'or', in its inclusive or exclusive sense, and concerned at most three entities in the domain of discourse. Participants nevertheless made systematically erroneous inferences.

Acknowledgments

This research was supported by a National Science Foundation Graduate Research Fellowship. 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 goes to Micah Clark for his critique of the model theory.

References

Bauer, M.I., and Johnson-Laird, P.N. (1993). How diagrams can improve reasoning. *Psychological Science*, 4.

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

Bucciarelli, M., & Johnson-Laird, P.N. (2005). Naïve deontics: A theory of meaning, representation, and reasoning. *Cognitive Psychology*, 50, 159-193.

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

García-Madruga, J.A., Moreno, S., Carriedo, N., Gutiérrez, F., & Johnson-Laird, P.N. (2001). Are conjunctive inferences easier than disjunctive inferences? A comparison of rules and models. *Quarterly Journal of Experimental Psychology*, 54A, 613-632.

Johnson-Laird, P.N. (2006). *How We Reason*. Oxford University Press.

Johson-Laird, P.N., & Byrne, R.M.J. (1991). *Deduction*. Psychology Press.

Johnson-Laird, P.N., Legrenzi, P., Girotto, V., & Legrenzi, M.S. (2000). Illusions in reasoning about consistency. *Science*, 288, 531-532.

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

Oaksford, M., and Chater, N. (1998). *Rationality in an Uncertain World*. Hove, UK: Psychology Press.

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

Rips, L.J. (1997). Goals for a Theory of Deduction: Reply to Johnson-Laird. *Minds and Machines*, 7, 409–424.

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

Appendix A

The table presents each sort of problem in Experiment 2, its mental models, its fully explicit models, the control and illusory questions with their correct answers, and the percentages of participants who drew them.

Problem:	Exclusive-exclusive: *(A and B) or else (B or else C)*		
Mental models:	A B B C	**Fully explicit models:**	A B C ¬A B ¬C A ¬B C ¬A ¬B C
		Correct answer	**% Correct**
Control:	Is it possible to have B only?	Yes	92
Illusion:	Is it possible to have A and B only?	No	26
Problem:	Exclusive-inclusive: *(A and B) or else (B or C)*		
Mental models:	A B B C B C	**Fully explicit models:**	¬A B ¬C A ¬B C ¬A ¬B C ¬A B C
		Correct answer	**% Correct**
Control:	Is it possible to have B only?	Yes	95
Illusion:	Is it possible to have A and B?	No	28
Problem:	Inclusive-exclusive: *(A and B) or (B or else C)*		
Mental models:	B C A B A B C	**Fully explicit models:**	A B C ¬A B ¬C A ¬B C ¬A ¬B C A B ¬C
		Correct answer	**% Correct**
Control:	Is it possible to have B and C only?	No	90
Illusion:	Is it possible to have A and C only?	Yes	38
Problem:	Inclusive-inclusive: *(A and B) or (B or C)*		
Mental models:	B C B C A B A B C	**Fully explicit models:**	¬A B ¬C A ¬B C ¬A ¬B C ¬A B C A B ¬C A B C
		Correct answer	**% Correct**
Control:	Is it possible to have neither A, B, nor C?	No	95
Illusion:	Is it possible to have A and C only?	Yes	56