Illusions in reasoning with quantifiers

Yingrui Yang (yingruiy@phoenix.princeton.edu)
Department of Psychology, Princeton University
Green Hall, Princeton, NJ 08544 USA

P.N. Johnson-Laird (phil@clarity.princeton.edu)
Department of Psychology, Princeton University
Green Hall, Princeton, NJ 08544 USA

Abstract

The mental model theory postulates that reasoners build models of the situations described in premises, and that these models normally make explicit only what is true. A computer program revealed an unexpected consequence of the theory: it predicts that certain inferences should have compelling but erroneous conclusions. Two experiments corroborated the existence of such illusions in inferences about what is possible given quantified assertions, such as 'At least some of the plastic beads are not red.' Experiment 1 showed that, as predicted, participants erroneously inferred that impossible assertions were possible, and that possible situations were impossible, but they performed well with control problems. Experiment 2 demonstrated the existence of similar illusions in inferences from dyadic assertions, e.g. 'All the boys played with the girls'.

Introduction

Consider 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?

Most people respond, 'yes'. The inference is a fallacy, because the conclusion would render both premises true, contrary to the rubric that only one of them is true. Such illusory inferences are predicted by the theory of mental models (Johnson-Laird and Byrne, 1991), though we have only recently discovered the prediction. Previous studies have corroborated the occurrence of illusions in inferences based on sentential connectives, such as "if" and "or" (see Johnson-Laird and Savary, 1996). A more powerful sort of reasoning, however, hinges on quantifiers, such as "some", "all", and "none", and so we examined illusory inferences based on quantified assertions. In this domain, the same premises can be used to elicit both illusory inferences and control inferences for which the theory predicts that the conclusions should be correct. In this way, we can eliminate the hypothesis that illusory premises are somehow too difficult for logically untrained individuals to understand. Our plan in what follows is to outline the theory of mental models for quantified assertions, and then to report two studies of illusory inferences in this domain.

Mental models of quantified assertions

How are quantified assertions represented in mental models? The most recent computer implementation (Bucciarelli and Johnson-Laird, 1998) postulates that an assertion of the form:

All the A are B

elicits a model with a small arbitrary number of tokens representing the relevant individuals, and that individuals make a mental footnote to indicate that the set of A's has been exhaustively represented, which we here represent by a square bracket:

```
[a]    b
[a]    b
  . . .
```

This model accordingly depicts two individuals who are A's and thus B's, and the ellipsis allows for the possibility of other sorts of individuals. The model exhausts the cases of individuals who are A's, i.e. individuals corresponding to the ellipsis cannot include individuals who are A's. Table 1 shows the models of the main sorts of singly-quantified assertion. Such models are isomorphic to one way of using Euler circles (cf. Stenning and Yule, 1997), but they generalize more readily to capture inferences hinging on relations, which cannot be captured within Euler circles, e.g.: All horses are animals. Therefore, All horses' heads are animals' heads. Likewise, human reasoners can construct alternative models of the premises (Bucciarelli and Johnson-Laird, 1998).

A fundamental principle of the model theory, the so-called 'principle of truth', is that reasoners normally represent only what is true in order to minimize the load on working memory. This principle is subtle because it applies at two levels: reasoners represent only true possibilities, and within those true possibilities they represent only those literal propositions (affirmative or negative) in the premises that are true. Likewise, the principle concerns falsity, which is not the same as

Table 1: The mental models of the four main singly-quantified assertions. The square brackets indicate that a set of individuals has been exhaustively represented, i.e. no more individuals of this sort can be added to the model. The initial model of Some A are not B supports the converse conclusion, but this conclusion can be refuted by an alternative model of the premise.

Assertions	Mental models	
All the A are B:	[a]	b
	[a]	b
	. . .	
Some of the A are B:	a	b
	a	
		b
	. . .	
None of the A is a B:	[a]	¬ b
	[a]	¬ b
		[b]
		[b]
	. . .	
Some of the A are not B:	a	¬ b
	a	¬ b
		b
		b

negation: negation is a syntactic notion, and a negative assertion can be true or false, whereas falsity is a semantic notion. We can best illustrate this principle with an example of sentential reasoning. Given an exclusive disjunction about a hand of cards, such as:

There is not a king in the hand, or else there is an ace in the hand

reasoners construct two alternative models, which we show here on separate lines:

¬ king
        ace

where '¬' denotes negation. Each model corresponds to a true possibility given the disjunction -- each model corresponds to a true row in a truth table, and each model represents only those literal propositions in the disjunction that are true within the possibility. Hence, the first model does not represent explicitly that it is false that there is an ace, and the second model does not represent explicitly that it is false that there is not a king. Reasoners make mental 'footnotes' to keep track of this false information, but these footnotes are soon likely to be forgotten, which in turn leads to the illusions. The footnotes make it possible to flesh out the models explicitly. And only fully explicit models of what is possible given the disjunction represent both the true and the false components in each model:

¬ king ¬ ace
king ace

According to the theory, however, reasoners do not normally construct fully explicit models, and retain the footnotes only for premises that tend to be simpler than the one above.

Illusory inferences with quantifiers

Consider the following problem again:

1. 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?

According to the model theory, reasoners should consider the model of the first premise:

p ¬ r
p ¬ r
    r
    r
  . . .

where 'p' denotes a plastic bead and 'r' denotes red. The model supports the possibility that None of the red beads are plastic, and so reasoners should respond, 'Yes'. They will make the same response if they examine the model of the second premise. In either case, however, the response is an illusion of possibility. When the first premise is true, the second premise is false, i.e. some of the plastic beads are red, and so the correct model is:

p r
p ¬ r
  . . .

Conversely, when the second premise is true, the first premise is false, i.e. all of the plastic beads are red, which conflicts with the second premise, and so the result is the empty (or null) model. The remaining model refutes the possibility that none of the red beads are plastic; and so the correct answer to the question is, 'No'. If we present the premises of problem 1 above, but ask a different question:

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

it should elicit the answer 'yes,' because the model based on the first premise can easily be modified to:

p ¬ r
p ¬ r
p r
    r
  . . .

Indeed, evidence suggests that many people construct this sort of model ab initio (Bucciarelli and Johnson-Laird, 1998). The answer 'yes' to this problem is correct, even though the procedure still fails to take into account what is false. Thus, the failure to consider falsity does not lead to errors with the control problems. With the premises of problem 1, the question:

3. Is it possible that all of the red beads are plastic?

should elicit an illusion of impossibility, i.e. reasoners should wrongly respond 'no' on the basis of their mental models. In contrast, the question:

4. Is it possible that all of the plastic beads are red?

should elicit the correct response, 'no'.

In Experiment 1, we gave 20 Princeton students inferences of each of four sorts: illusions of possibility and their controls, and illusions of impossibility and their controls. The inferences were based on five pairs of monadic premises that each referred to the same two terms. The five pairs of premises were combined on separate trials with four different modal conclusions. Ten of the problems were illusions, and ten of them were controls. The 20 problems were presented in one of six different random orders to each participant, and each problem was based on a different set of contents. The participants rated their confidence in each of their answers on a five-point scale.

Table 2: The percentages of correct responses to the four sorts of problems in Experiment 1. The figures in parentheses are the participants' mean confidence in their answers (on a five-point scale, where 1 signifies 'No confidence' and 5 signifies 'Full confidence').

	Illusions	Controls
Inferences of possibility	23 (4.35)	87 (4.43)
Inferences of impossibility	67 (4.36)	94 (4.71)
Overall	54 (4.39)	90 (4.51)

Table 2 presents the results. The participants were correct on 90% of the control inferences, but were only slightly better than chance, 54%, with the illusory inferences. All 20 of the participants were more accurate with the control inferences than with the illusory ones ($p = .5^{20}$, i.e. less than 1 in a million). The participants were also slightly more confident in their answers to the control problems than to the illusory problems, and this difference -- a mean of .13 on the five-point scale (Wilcoxon's test, $N = 12$, $T+ = 65$, $p < .03$). There was a reliable interaction: the difference in accuracy between the control problems and the illusions was greater for the inferences of possibility than for the inferences of impossibility (Wilcoxon's test, $z = 2.77$, $p < .01$).

The results corroborated the model theory, but the illusions of possibility were more telling than those of impossibility. To infer that a situation is impossible, however, calls for a check of every model, whereas to infer that a situation is possible does not, and so reasoners are less likely to make the inference of impossibility. This difference occurs in harder problems that are not illusory (Bell and Johnson-Laird, 1998).

Experiment 2 was a replication of the previous experiment, but using a new set of inferences in which the premises concerned dyadic relations, e.g. 'All the boys got red beads'. In these problems, the premises contained a main verb that described a relation, such as 'got', that holds between two sets of individuals. Twenty Princeton students carried out four sorts of inferences: illusions of possibility, their controls with the correct answer 'yes', illusions of impossibility, and their controls with the correct answer 'no'. The design and procedure were the same as those of Experiment 1.

Table 3: The percentages of correct responses to the four sorts of problems in Experiment 2. The figures in parentheses are the participants' mean confidence in their answers (on a five-point scale).

	Illusions	Controls
Inferences of possibility	13 (4.37)	94 (4.31)
Inferences of impossibility	70 (4.12)	93 (4.69)
Overall	50 (4.22)	93 (4.54)

Table 3 presents the results, which were very similar to those of the previous experiment. All 20 of the participants were more accurate with the control inferences than with the illusory ones ($p = .5^{20}$, i.e. less than 1 in a million). They were also slightly more confident in their answers to the control problems than to the illusory problems (Wilcoxon's test, $z = 3.62$, $p < .001$). And, once more, the difference between the control problems and the illusions was greater for the inferences of possibility than for the inferences of impossibility (Wilcoxon's test, $z = 3.22$, $p < .001$). The correlation between the results of the two experiments was very high (Kendall's tau = .94, $p < .0001$). Evidently, the pattern of results with these quantified inferences is highly robust.

General Discussion

The experiments showed that individuals succumb to illusions in inferences about what is possible, and what is impossible, given quantified assertions about individuals. The phenomena were predicted by the model theory's principle of truth according to which reasoners take into account what is true, but not what is false. Is there an alternative explanation? One hypothesis is that the instructions, materials, or task, were too difficult for the participants to understand. Two phenomena, however, count against this idea. First, the participants performed very well with the control problems (over 90% correct in each experiment). These control problems were based on logically identical premises to those used for the illusions, and they differed only in the question posed to the participants. These questions themselves varied from one problem to another, so that, for example, a question of the form: 'Is it possible that none of the A are B?' occurred sometimes for an illusion and sometimes for a control problem. Hence, there is unlikely to be anything about the instructions, materials, or task, that is intrinsically difficult to understand. Second, the participants were confident in their responses, albeit slightly more confident in their responses to the controls than to the illusions.

Another hypothesis is that the participants made their inferences based on whether or not there was an exact match between the questioned conclusion and one of the premises. For example, they responded 'yes' when the questioned conclusion matched a premise. But, if the participants had responded 'yes' whenever there was a match, and 'no' otherwise, then the pattern of their responses would have been wholly different from those that we observed, e.g. the participants would have responded 'no' uniformly to problems 1 through 4 above.

In two recent experiments, we have further corroborated the model theory's account of the quantified illusions. The experiments used the same sorts of inferences as Experiment 1, but after half the problems, the participants received some remedial instructions. When they were told to check that their putative conclusions were consistent with the truth of the first premise and the falsity of the second premise, there was a significant improvement in their performance with the illusory problems. Likewise, when they were told to check that their putative conclusions were consistent with both relevant cases -- 1) with the truth of the first premise and the falsity of the second premise, and 2) with the truth of the second premise and the falsity of the first premise -- the gap between the illusions and the controls disappeared. There was both a significant improvement with the illusions and a significant decline with the controls, and performance converged at 71% correct. The decline with the control problems is presumably attributable to the need to work through a costly procedure that is unnecessary with problems that the participants tend to get right even when they neglect falsity.

The illusions are predicted by the model theory but they are not readily explained by theories based on formal rules of inference (see e.g. Braine, 1998). These theories apply valid rules of inference to the logical forms of premises, and so they cannot predict the systematic occurrence of invalid conclusions. Is it possible to save the formal rule theories? One solution is trivial. Formal rules, in the broadest sense, have universal Turing machine power, and so they are equivalent to a programming language (Rips, 1994). Hence, they can be used to simulate any theory whatsoever, including the model theory. It is this principle, of course, which we exploited in implementing the model theory in a computer program. In this broad sense, formal rules are no more refutable than a programming language. In the narrower sense of the current theories, their rules contain no way to embody the principle of truth. Perhaps a more plausible idea to save the current theories is that reasoners somehow misapply a suppositional strategy (Luca Bonatti, David O'Brien, personal communications). Recent empirical studies of Braine's theory (Yang, Braine, and O'Brien, 1998) have investigated only direct reasoning with quantified assertions. But, in the case of problem 1 above, for example, a supposition of the first premise, <u>At least some A are not B</u>, fails to yield the conclusion <u>No B are A</u>. But, a supposition of the second premise, <u>No A are B</u>, yields the conclusion <u>No B are A</u>. This inference, in turn, leads to the response 'yes' to the question. But, now consider problem 2, which has the same premises, but a different question: 'Is it possible that some B are A?' The putative conclusion cannot be inferred from a supposition of either premise, and so reasoners should respond 'no' to the question. In fact, the vast majority of participants in both experiments <u>correctly</u> responded, 'yes', to this control problem. In other words, a misapplication of suppositions here obliterates the distinction between illusory and control problems. Hence, an appropriate simple modification to current formal rule theories may not exist.

The results of our experiments corroborate the mental model theory. Other studies of illusions in sentential and probabilistic reasoning have the same moral (Johnson-Laird and Savary, 1996; Johnson-Laird and Goldvarg, 1997). What the illusions have in common is the failure to consider falsity at either of two levels: a neglect of possibilities that are false, and a neglect within true possibilities of those literal propositions (affirmative or negative) in the premises that are false. The principle of truth is therefore highly plausible:

logically-untrained individuals tend to reason on the basis of what is true, not what is false. With hindsight, other well-known phenomena in reasoning appear to be further manifestations of the principle of truth, e.g. the difficulty of Wason's selection task, and the difficulty of inferences in the form known as modus tollens (If A then B, not-B, ∴ not-A).

Formal rules and mental models do not exhaust the domain of possible theories of reasoning. Neither sort of theory tells the whole story about reasoning, and some new theory may well transcend them. Our results suggest that this new theory will have to accommodate the principle of truth.

Acknowledgments

We are grateful to colleagues for their helpful advice: Patricia Barres, Victoria Bell, Monica Bucciarelli Zachary Estes, Yevgeniya Goldvarg, and Mary Newsome. The research was supported in part by the Xian-Lin Ji Foundation of Peking University, China.

References

Bell, V.A., and Johnson-Laird, P.N. (1998) A model theory of modal reasoning. Cognitive Science, in press.

Braine, M.D.S. (1998) Steps towards a mental predicate logic. In Braine, M.D.S., and O'Brien, D.P. (Eds.) Mental Logic. Mahwah, NJ: Lawrence Erlbaum Associates.

Bucciarelli, M., and Johnson-Laird, P. N. (1998) Strategies in syllogistic reasoning. Under submission.

Johnson-Laird, P. N., and Byrne, R. M. J. (1991). Deduction. Hillsdale, NJ: Lawrence Erlbaum Association.

Johnson-Laird, P. N., and Goldvarg, Y. (1997). How to make the impossible seem possible. Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society. 354-357.

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

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

Stenning, K., and Yule, P. (1997) Image and language in human reasoning: A syllogistic illustration. Cognitive Psychology, 34, 109-159.

Yang, Y., Braine, M.D.S., and O'Brien, D.P. (1998). Some empirical justifications of the mental predicate logic model. In Braine, M.D.S., and O'Brien, D.P. (Eds.) Mental Logic. Mahwah, NJ: Lawrence Erlbaum Associates.