Problem representations and illusions in reasoning

Vladimir M. Sloutsky (sloutsky.1@osu.edu)
Center for Cognitive Science & School of Teaching & Learning; 1945 N. High Street
Columbus, OH 43210, USA

Philip N. Johsnon-Laird (phil@clarity.princeton.edu)
Department of Psychology, Princeton Univeristy
Green Hall, Princeton, NJ 08554 USA

Abstract

The mental model theory of reasoning postulates that reasoners build models of the situations described in premises, and that these models normally make explicit only what is true. The theory has an unexpected consequence: it predicts the occurrence of inferences that are systematically invalid. These inferences should arise from reasoners failing to take into account what is false. We report an experiment that corroborated the occurrence of these illusory inferences, and that eliminated a number of alternative explanations for them. Results illuminate the controversy among various current theories of reasoning.

Introduction

While reasoning is ubiquitous in human life, its underlying mechanisms are a matter of controversy. Although there are many theories of reasoning with conditionals (e.g., Cheng & Holyoak, 1985; Cosmides, 1989), there are two major general approaches to reasoning: the syntactic and the semantic. According to the syntactic approach, reasoning hinges on a set of formal rules of inference or "inferential schemata" (Braine & O'Brien, 1998; Rips, 1994). For instance, when individuals are presented with premises of the following form:

A or *B*
not-*A*

they draw the conclusion:

B

using the "disjunction elimination schema", which is a formal rule that has precisely the form of this inference.

According to the semantic approach, the untrained mind is not equipped with formal rules of inference, but relies instead on the validity of arguments. The principle states that an inference is valid if its conclusion must be true given that its premises are true (e.g. Johnson-Laird & Byrne, 1991; Polk & Newell, 1995). Among other theories within the semantic approach, the mental logic theory (Johnson-Laird & Byrne, 1991) attempts to explain both correct and erroneous reasoning using a small number of fundamental assumptions. These include: (1) people represent information in premises in a systematic manner and (2) these representations are constructed in accordance with a principle of truth:

- Individuals normally represent only true possibilities.
- Within these possibilities they represent only those literal propositions in the premises (affirmative or negative) that are true.

Given the example above, reasoners therefore construct the following mental models, where each row denotes a mental model of a possibility, and "¬" denotes negation:

First premise	Second premise
A	¬A
B	
A B	

They can infer that *B* is the case, because the second premise rules out those models of the first premise in which *A* occurs, i.e., the first and third models of possibilities.

Although the syntactic and semantic approaches postulate different mechanisms underlying reasoning, they often yield similar predictions. Yet, there are inferences for which their predictions diverge. In particular, the principle of truth leads to the prediction that certain inferences should lead reasoners systematically astray. For example, consider the following inference in which you must assume that only one of the two conditional premises is true:

If George was on the team then Dan was on the team.
If Dan wasn't on the team than George was on the team.
Assuming that only one of the above statements is true, is it possible that George was on the team, and Dan was on the team?

The mental model theory predicts that reasoners should envisage the following mental models of possibilities for the first premise:

George Dan
…

where the first model makes explicit the possibility in which both George and Dan are on the team, and the second model (ellipses) corresponds to those possibilities in which the antecedent of the conditional is false. The theory accordingly assumes that individuals do not normally make these possibilities explicit (Johnson-Laird & Byrne, 1991). Likewise, reasoners should envisage the following mental models for the second premise:

¬ Dan George
…

According to the principle of truth, when reasoners think about the truth of one premise, they will fail to think about the falsity of the other premise. The question posed in the problem is whether it is possible that George was on the team and Dan was on the team. This situation corresponds to a possibility in the mental models of the first premise, and so reasoners should respond “yes”. The response is an illusion, however. Indeed, the fully explicit models of the premises are respectively:

First premise		Second premise	
George	Dan	¬ Dan	George
¬ George	Dan	Dan	George
¬ George	¬Dan	¬ Dan	¬George

It follows that the question *Is it possible that George was on the team, and Dan was on the team* has the correct answer, "no", because the case in which both George and Dan are members of the team is true in both premises. We label this inference as a "yes/no" problem, where the first word ("yes") is the predicted answer and the second word ("no") is the correct answer. The same premises support a control inference to which reasoners should get the correct answer even though they fail to represent what is false: *Is it possible that George was on the team, but Dan wasn't on the team?* The situation corresponds to a mental model of the second premise, but it is also correct because it occurs only in the fully explicit models of the second premise. We define such control inferences as "yes/yes" problems.

By pairing the same premises with two other sorts of question, we can create inferences to which reasoners should fall into the trap of responding "no" incorrectly ("no/yes" problems): *Is it possible that George was not on the team, and Dan was not on the team?* We can also create a corresponding control problem to which reasoners should respond "no" correctly ("no/no" problems): *Is it possible that George wasn't on the team, but Dan was on the team?*

The source of the illusions according to the model theory is the failure to represent what is false. This failure is likely to be more pronounced among those who have little or no training in reasoning, and among those who score less well on SAT tests, which Keith Stanovich (personal communication) has shown to correlate significantly with logical performance.

It is also known from previous work on reasoning about possibilities (Bell & Johnson-Laird, 1998) that it is easier for people to establish that a conclusion is possible than it is to establish that the conclusion is impossible. However, it is easier for them to establish that a conclusion is not necessary than that it is necessary. The reason for this divergence is that to establish that a conclusion is impossible or necessary, one needs to search through all models, whereas to establish that a conclusion is possible or unnecessary, it is sufficient to find just one example or counterexample. Therefore, we can predict that both illusory and control problems that require participants to answer "No" (impossible), should be harder than respective problems requiring them to answer "Yes" (possible). As a result, Yes/No illusion should be harder than No/Yes illusions, and No/No controls should be harder than Yes/Yes controls.

In contrast to the model theory, current theories based on formal rules of inference (e.g. Rips, 1994; Braine & O'Brien, 1998) do not predict the occurrence of the illusory inferences. These theories rely solely on valid rules of inference, and so in principle they cannot predict the occurrence of systematically erroneous conclusions. Because both illusory and control problems are based on the same premises, the formal rules theorist should predict no systematic differences between illusory and control problems.

An experiment comparing illusory inferences and control inferences

In order to test the model theory’s predictions, we carried out an experiment in which we examined performance on four types of inference problems. They were illusory inferences with a predicted response of "yes" (yes/no problems), illusory inferences with a predicted response of "no" (no/yes problems), and their respective controls (yes/yes problems) and (no/no problems). The four types of problems were based on the same premises so that no differences in the premises could be responsible for the results. Similarly, all the questions following the premises were in the form of conjunctions. The form of the problems is summarized in Table 1. The problems in set A are those that we described in the Introduction. Those in set B are comparable with the four types of problems all based on the same premises.

Method

Participants Two groups of undergraduate students participated in the experiment. One group of 18 were recruited from a private (highly selective) university (*M* = 20.6 years, *SD* = 1.4; 11 men and 7 women) and 20 from a large public (mainly non-selective) university (*M* = 20.0 years, *SD* = 1.5; 9 men and 11 women). Hence, the two groups were drawn from two populations, which differ in their required SAT scores and in the emphasis in their curricula on mathematical training.

Materials Each participant carried out 16 problems (the eight sorts of problems in Table 1 and eight filler items) in one of two random orders. The content of the problems concerned team memberships and each of the 16 problems was about a different pair of individuals, i.e. frequent one- or two-syllable first names of males and females.

Table 1: The inferences in the experiment, their mental models, and their fully explicit models. In each set, only one of the premises is true.

Problem set A	Mental models		Fully explicit models	
	Premise 1	Premise 2	Premise 1	Premise 2
If B then A If not A then B	B A …	¬A B …	B A ¬B A ¬B ¬A	¬A B A ¬B A B
1. Is B & A possible? (yes/no)	B & A are in the first premise		B & A are in both premises	
2. Is B & not-A possible? (yes/yes)	B & ¬ A are in the second premise		B & ¬A are in the second premise	
3. Is not B & not-A possible? (no/yes)	¬B & ¬A are not in the premises		¬B & ¬A are in the first premise	
4. Is not B & A possible? (no/no)	¬B & A are not in the premises		¬B & A are not in the premises	
Problem set B	Mental models		Fully explicit models	
	Premise 1	Premise 2	Premise 1	Premise 2
If B then not A If A then B	B ¬A …	A B …	B ¬A ¬B A ¬B ¬A	A B ¬A B ¬A ¬B
1. Is B & not A Possible? (yes/no)	B & ¬A are in the first premise		B & ¬A are in both premises	
2. Is A & B Possible? (yes/yes)	A & B are in the second premise		A & B are in the second premise	
3. Is not B & A possible? (no/yes)	B & ¬A are not in the premises		¬B & A are in the first premise	
4. Is not A & not B possible? (no/no)	¬A & ¬B are not in the premises		¬A & ¬B are in both premises	

Procedure The participants were tested individually. The experimenter read them the instructions and presented them with a warm-up problem. The key component of the instructions was as follows: *Imagine that there is a meeting of two old coaches who coached together two competing teams, the Bulls and the Wildcats. They started talking about the good old days when their teams competed. But it soon turned out that as in the good old days they could not agree on anything. In particular, they weren't even able to agree on who was on each team. So they might need your help. But before helping them, I want you to know that in every argued case, they cannot be both right or both wrong. In other words, in each case one of them is right and the other is wrong. Sometimes it is the coach of the Bulls (Coach Bull) who is right, sometimes it is the coach of the Wildcats (Coach Wildcat), but it is always the case that only one is right and another is wrong. Your goal is to decide whether or not some of the things they say are possible, given that one is right and another is wrong.*

The participants then carried out a simple warm up problem:

Coach Bull: *Sara wasn't on the team*

Coach Wildcat: *Megan wasn't on the team*

Is it possible that Sara and Megan were on the team? Why?

If a participant gave an incorrect answer (i.e., if they asserted that it was possible for Sara and Megan to be on the team), the experimenter provided an explanation and offered another warm-up problem. The experimenter also took special care to explain to the participants that the coaches were always arguing about the same team. Once the participants had understood the instruction, they proceeded to the experiment proper, which lasted for approximately half an hour. The problems were in booklets, and the participants were told to respond “yes” or “no” to each question.

Results

Table 2 presents the results of the experiment for the two groups of participants. The table shows that both groups of participants tended to succumb to the illusions (only 31% correct Yes/No inferences and 50% correct No/Yes inferences), but performed better on the control problems (79% correct Yes/Yes inferences and 66% correct No/No inferences). Across groups, participants performed significantly better on control problems than on illusory problems (Wilcoxon z = 3.59, p<.0001).

In the Table, the percentages for the group of the selective university students are on the left of each cell, and those for the non-selective students are in parentheses on the right of each cell. All eight percentages depart significantly from chance (p ranging from < .0001 to <

.05).

Table 2: The percentages of correct responses to the four sorts of problems for the two groups of participants.

Illusory inferences		Control inferences	
1. Yes/No	33 (27)	2. Yes/Yes	86 (73)
3. No/Yes	67 (35)	4. No/No	70 (65)

Note that students in the non-selective university were more likely to succumb to illusions, and their results conformed better to the mental model theory's predictions than did results of students in the selective university. Both groups responded "yes" to the Yes/No illusions significantly more often than a chance rate (all *ps* < .05 or better), but only the public university students succumbed to the "No/Yes" illusions significantly more often than a chance rate ($p <$ .05). Both groups performed better than chance on the control problems (*p*s ranged from < .0001 to < .05), and there were no differences on these problems between the selective university the non-selective university students. At the same time, the two groups differed on Yes/No (Mann-Whitney U = 190, *p*<. 05) and No/Yes (No (Mann-Whitney U = 255, *p*<. 03) illusory problems. Finally, as predicted the yes/no illusory problems were harder than the no/yes illusions, Wilcoxon z = 2.1, p < .05, whereas the yes/yes controls were somewhat easier than the no/no controls, Wilcoxon z =1.8, p =.07.

Discussion

A surprising consequence of the mental model theory of reasoning is its prediction of illusory inferences, that is, inferences that lead to systematic but fallacious conclusions. The results of the experiment corroborated this prediction of the theory. For example, given a problem of the form below, most of our participants wrongly concluded that the answer was "yes".

Assuming that only one of the above statements is true:

If A then B.
If ¬ B then A.
Is A & B possible?

Is there any plausible alternative explanation of our results apart from the failure to represent what is false? One alternative hypothesis is that the task, instructions, or premises of the inferences are so complex, ambiguous, or pragmatically odd, that they confused the experimental participants, thus adversely affecting their performance. However, this hypothesis fails to explain the systematicity and predictability of errors, as well as the correct performance of the control problems. Another alternative hypothesis is that the participants failed to notice that one premise is true and the other premise is false. Again, this idea is most likely implausible given the framing of the problems as disagreements between two coaches, and the participants' practice with such a problem.

A more plausible alternative hypothesis is that that conditionals have an interpretation distinct from the one that we have proposed. There are various versions of this hypothesis, e.g., conditionals are interpreted as having a "defective" truth table (e.g. Wason & Johnson-Laird, 1972), or as having some other, as yet unknown, sophisticated meaning. The hypothesis of a "defective" truth table treats conditionals as having no truth value whenever their antecedents are false. The idea, however, runs into insuperable difficulties with biconditionals, such as*:*

If, and only if, Dan was in the game, then George was in the game.

People judge that this assertion is true when both Dan and George were in the game or not in the game; but when one was in the game and the other was not, they judge it to be false. Hence, the biconditional has a complete truth table. Yet, it can be paraphrased by the following conjunction of two conditionals:

If Dan was in the game then George was in the game, and if George was in the game then Dan was in the game.

Consider the case where neither Dan nor George was in the game. Neither of the two conditionals has a truth value, yet the biconditional is true. How can the conjunction of two assertions lacking a truth value yield an assertion that is true? The answer is: it cannot.

As for an unknown sophisticated meaning, our analysis offers a parsimonious explanations that rests only on two simple and testable assumptions, such as (1) the mental model representation of the conditional and (2) the principle of truth. Recall that the first states that people construe only two models to represent the conditional, such as *If A then B*:

A B
…

and the second states that people represent only true possibilities.

Robert Mackiewicz and Walter Schaeken (personal communication) have drawn our attention to an interesting possibility. Perhaps, if reasoners think about one proposition in a disjunction then they forget about the other. However, if reasoners merely forgot one of the clauses, then they should be liable to forget A or to forget B in dealing with a conjunction of A and B (cf. Johnson-Laird & Savary, 1996). However, neither adults nor children usually forget the constituents of conjunctions when they reason (Sloutsky, Rader, & Morris, 1998; Morris & Sloutsky, 1999). On the other hand, for exclusive disjunctions, such as the one in our experiment, it seems that people think about the truth of one proposition while forgetting about the falsity of the other proposition. Such dissociation between not forgetting propositions in conjunctions and forgetting them in exclusive disjunctions, if demonstrated empirically, would strongly support the principle of truth.

Unlike these rival hypotheses, the mental model theory of reasoning predicts that any manipulation that emphasizes falsity should reduce the illusions. Recent studies have corroborated this prediction. For example, the rubric, "Only one of the following two premises is false," reliably reduced the illusions (Tabossi, Bell, & Johnson-Laird, 1998). They were reduced when the participants had to generate false instances of the premises before they carried out the inferential task (Newsome & Johnson-Laird, 1996). Likewise, they were reduced when the participants had to check whether the conclusions were consistent with the relations between the premises (Goldvarg & Johnson-Laird, 1999; Yang & Johnson-Laird, 1998a; 1998b). A final advantage of the model theory is that it bases its predictions on a single principle -- reasoners take into account truth, not falsity.

There is, however, another potential source of difficulty. Reasoners cannot cope very well with inferences that call for multiple models of the premises. Indeed, they often appear to construct only a single model. Bauer & Johnson-Laird (1993) reported that the most frequent error in a study of disjunctive inferences was that the participants constructed only one model of the premises. Likewise, children often appear to construct only a single model. Hence, a more radical source of error than the principle of truth is that reasoners may sometimes construct a “minimalist” representation of just a single possibility – just a single mental model of the premises (Sloutsky, Rader, & Morris, 1998; Sloutsky & Goldvarg, 1999).

The illusions are an important shortcoming in human reasoning, and they are worth investigating further for their own intrinsic interest. The neglect of falsity, however, appears to underlie a number of other well-established inferential phenomena, such as Wason’s selection task and the difficulty of *modus tollens* inferences (for a review, see Evans, Newstead, & Byrne, 1993). The illusions contravene all current formal rule theories (e.g. Braine & O’Brien, 1998; Rips, 1994). These theories rely solely on valid rules of inference, and so the only systematic conclusions that they can account for are valid ones. These theories therefore need to be amended -- either in their implementation or in a more radical way in order to account for the illusions. Our study shows that naï ve reasoning depends crucially on how individuals represent the premises.

References

Bauer, M. I., & Johnson-Laird, P. N. (1993). How diagrams can improve reasoning. *Psychological Science*, *4*, 372-378.

Bell, V. A., & Johnson-Laird, P. N. (1998). A model theory of modal reasoning. *Cognitive Science*, *1*(22), 25-51.

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

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

Cosmides, L. (1989) The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition*, *31*, 187-276.

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

Goldvarg, Y., & Johnson-Laird, P. N. (1999). *Memory & Cognition*, in press.

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.

Morris, B. J., & Sloutsky, V. M. (1998). Developmental differences in young children’s solutions of logical vs. empirical problems. In *Proceedings of the Twenty First Annual Conference of the Cognitive Science Society*. (pp. 000-000). Mahwah, NJ: Erlbaum.

Newsome, M. R., & Johnson-Laird, P. N. (1996). An antidote to illusory inferences? In Cottrell, G.W. (Ed.) *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates, p. 820.

Oaksford, M., & Stenning, K. (1992). Reasoning with conditionals containing negated constituents. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *18*, 835-854.

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

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

Sloutsky, V. M., & Goldvarg, Y. (1999). Effects of externalization on representation and recall of indeterminate problems. In *Proceedings of the Twenty First Annual Conference of the Cognitive Science Society*. (pp. 000-000). Mahwah, NJ: Erlbaum.

Sloutsky, V. M., Rader, A., & Morris, B. (1998). Increasing informativeness and reducing ambiguities: Adaptive strategies in human information processing. In Gernsbacher, M.A., & Derry, S.J. (Eds.) *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society*. (pp. 997-999). Mahwah, NJ: Erlbaum.

Tabossi, P., Bell, V. A., & Johnson-Laird, P. N. (1998). Mental models in deductive, modal, and probabilistic reasoning. In Habel, C., & Rickheit, G. (Eds.) *Mental Models in Discourse Processing & Reasoning*. Amsterdam: North-Holl&, in press.

Wason, P. C., & Johnson-Laird, P. N. (1972). *The Psychology of Deduction*: *Structure & Content*. Cambridge, MA: Harvard University Press. London: Batsford.

Yang, Y., & Johnson-Laird, P. N. (1998a). Illusions in quantified reasoning: How to make the impossible seem possible, and *vice versa*. *Memory & Cognition*, in press.

Yang, Y., & Johnson-Laird, P. N. (1998b) Systematic fallacies in quantified reasoning and how to eliminate them. Under submission.

Acknowledgements

This research has been supported by a grant from the James S. McDonnell Foundation awarded to the first author.