ELSEVIER

COGNITION

Cognition 76 (2000) 175–178

www.elsevier.com/locate/cognit

Discussion

Illusions and models: a reply to Barrouillet and Lecas

P.N. Johnson-Laird*

Department of Psychology, Princeton University, Princeton, NJ 08544, USA

Consider the following problem about a particular hand of cards:

If there is a king in the hand then there is an ace or else if there is not a king in the hand then there is an ace.

There is a king in the hand.

What follows?

Most people, experts and novices alike, infer that there is an ace. This inference is invalid, however. It is an example of an illusory inference, and such inferences – of which there are many other sorts – were predicted by the theory of mental models (Johnson-Laird and Savary, 1999). The prediction derives from the theory's principle of *truth*, which postulates that mental models represent true possibilities and that within each such possibility they represent clauses in the premises only when they are true. Barrouillet and Lecas (2000) report two important new results about illusory inferences. The aim of this reply is to offer a defense of the original version of the model theory.

(1) Barrouillet and Lecas claim that the original version of the model theory proposed the following mental models for the disjunction above, where ' ¬ ' denotes negation:

king ace

…

¬ king ace

…

and not the models that Johnson-Laird and Savary (1999) state in order to predict the illusions:

* Tel.: +1-609-258-4432; fax: +1-609-258-1113.
E-mail address: phil@princeton.edu (P.N. Johnson-Laird).

0010-0277/00/$ - see front matter © 2000 Elsevier Science B.V. All rights reserved.
PII: S0010-0277(00)00076-7

king ace
¬ king ace
…

In fact, the computer program implementing the model theory, as briefly described in Johnson-Laird and Savary, *does* construct the preceding models. When it works at its most primitive level, there is no need to have two instances of the same implicit model (as represented by the ellipsis). When the program operates at an advanced level, however, it constructs the following fully explicit models, which are correct:

king ¬ ace
¬ king ¬ ace

where ' ¬ ' signifies negation. The author noticed this discrepancy in the outputs of the two levels, and inferred that there was a bug in the program. To his chagrin, he discovered that the bug was in his mind. The program was right; he had succumbed to an illusory inference. The model theory as implemented in the program accordingly predicted the illusory inferences, and it produced the models of the disjunction of the two conditionals that are also the models of the following assertion:

If there is a king or there is not a king then there is an ace.

(2) Barrouillet and Lecas's Experiment 1 used materials in which it was clear that the disjunction applied to the conditionals as a whole. They introduced two fictitious people, who respectively assert the two conditionals, and they told the subjects that one person said something true and one person said something false. This manipulation reduced the frequency of illusory inferences in comparison with a control condition, but it did not eliminate them. The result is important. However, it is consistent with the model theory. The manipulation should make it easier for the participants to envisage what is false. Robert Mackiewicz and Walter Schaeken (pers. commun.) have likewise obtained enhanced effects by emphasizing that one of two separate speakers is known to be unreliable.

(3) Barrouillet and Lecas's Experiment 2 showed that subjects find it harder to infer what follows from the falsity of a conditional than from the falsity of a disjunction. They argue that their version of the model theory predicts the difference, whereas the original version of the theory does not. It is not clear, however, that their version of the theory does predict the difference. They write: "our theory of conditional reasoning…predicts that the negation of an 'if then' sentence should be more difficult to calculate than the negation of an 'or' sentence because the models for the conditional are hypothetical and differ in nature from those for the disjunction…" But why should it be harder to negate hypothetical models that differ in nature from models of disjunctions? Moreover, the process of envisaging false instances may be different in English than in French. Several studies of English have failed to detect that it is harder to envisage false conditionals than false disjunctions (Barres & Johnson-Laird, 1997; Sloutsky & Goldvarg, 2000). Similarly, Newsome and Johnson-Laird (1996) in a therapeutic study of the illusions found that the subjects had no

difficulty in stating the false instances of conditionals. Barrouillet and Lecas's result accordingly calls for a replication in English, and further studies to examine whether part of the difficulty of illusions based on conditionals depends on drawing conclusions from their falsity.

(4) Barrouillet and Lecas's theory of mental models postulates that the models of conditionals are 'relational and hypothetical'. The key issue here is to determine what is at stake semantically. Consider the fully explicit models of a conditional, *If A then B*, according to the original theory of mental models:

a b

¬ a b

¬ a ¬ b

They show a relation between A and B, namely, A does not occur without B. Likewise, since each model represents a possibility, the model containing A is hypothetical, that is, A may or may not occur. On this account, the conditional can be paraphrased as: *If A then B, and if not A then B may or may not be the case*. The following simple inference is valid and likely to be accepted by most people:

A or B, or both.

Therefore, if not A then B.

Hence, the model of the disjunction in which A does not occur is presumably hypothetical. Indeed, it is arguable that whenever an assertion has more than one model, each individual model represents a hypothetical possibility.

Is anything more at stake in the semantics of conditionals? My surmise is that Barrouillet and Lecas would argue that the present account is incomplete. They write: "The values of the consequent [of a conditional] are relevant only if the hypothesis they are linked to holds." But, there are counterexamples to this claim. Consider, for example, the following remark, which the present author has been known to make:

If that experiment works then I'll jump into Lake Carnegie.

According to Johnson-Laird and Byrne (2000), the obvious falsity of the consequent rules out two of the otherwise feasible models of the conditional to leave only the following model:

¬ works ¬ jumps

So, the hypothesis to which the consequent is linked does not hold. Similarly, one can assert:

If Bill Gates needs money, which he doesn't, then I'll be happy to lend him some.

The consequent is true, and relevant, even though the antecedent is explicitly false. No doubt many conditionals are interpreted as conveying a particular sort of relation between their antecedents and consequents. But, as Johnson-Laird and Byrne argue, their *meaning* alone does not signify any such relation. If it did, then to deny the

relation whilst asserting the conditional would be to contradict oneself. Yet, the next example makes good sense:

If there was a circle on the board then there was a star on the board, though there was no relation between the two – they merely happened to co-occur.

(5) Barrouillet and Lecas draw the following moral: "it is not clear that (a) the illusory inferences resulting from a disjunction of conditionals…result from the principle of truth, and (b) that the standard mental models theory can account for this phenomenon." They may well be right that more is at stake than the principle of truth in illusions based on conditionals. Readers new to illusory inferences, however, should note that there are plenty of other robust illusions that do not depend on disjunctions of conditionals (see Goldvarg & Johnson-Laird, 2000; Johnson-Laird & Savary, 1996, 1999; Johnson-Laird, Legrenzi, Girotto, Legrenzi & Caverni, 2000; Yang & Johnson-Laird, 2000). Like the conditional illusions, however, these illusions were also predicted by the computer program implementing the principle of truth.

References

Barres, P. E., & Johnson-Laird, P. N. (1997). Why is it hard to imagine what is false? *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society*, p. 859.

Barrouillet, P., & Lecas, J.-F. (2000). Illusory inferences from a disjunction of conditionals: a new mental models account. *Cognition, 76, 3–9.*

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

Johnson-Laird, P. N., & Byrne, R. M. J. (2000). Conditionals: a theory of meaning, pragmatics, and inference. Manuscript submitted for publication.

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

Johnson-Laird, P. N., Legrenzi, P., Girotto, V., Legrenzi, M., & Caverni, J. -P. (in press). Illusions in reasoning about consistency. *Science*.

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

Sloutsky, V. M. & Goldvarg, Y. (2000). Representation and recall of determinate and indeterminate problems. Manuscript submitted for publication.

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