Authors' Response

Models, necessity, and the search for counterexamples

P. N. Johnson-Laird* and Ruth M. J. Byrne
*Department of Psychology, Princeton University, Princeton, NJ 08544;  
**Department of Psychology, Trinity College, University of Dublin, Dublin 2.  
Contact: phil@clarity.princeton.edu

The latest commentators on the mental model theory (Johnson-Laird & Byrne 1991; 1993) take different positions about its central assumption. Martin-Cordero & González-Labra accept that a psychological theory of reasoning should be semantic, and that the theory of mental models shows how deductions can be made without relying on formal rules of inference. We are grateful to them for their broad approval of the theory. What worries them is its lack of any strong principles about how general knowledge is mentally represented, or about how reasoners search for models that are counterexamples to their conclusions. In contrast, Smith does not accept a purely semantic theory of deduction and doubts whether the model theory can give a satisfactory account of the necessity of valid conclusions. We are grateful to him for giving us an opportunity to demonstrate how the theory accounts for necessity. We will consider first the idea of supplementing the model theory with formal principles, then the search for counterexamples, and finally the representation of knowledge.

R1. Models and necessity

A deductively valid conclusion is necessarily true given the truth of its premises. It holds in all possibilities compatible with the premises. To establish validity, reasoners must accordingly ensure that there is no model of the premises in which the conclusion is false. Smith refers to the "inductive" character of the search for such counterexamples, which, he claims, leads to the inability of the theory to cope with necessity. He believes that necessity can be accommodated only by the addition of formal rules of inference to the model theory. "There are grounds," he writes, "for taking Piaget to be an 'arch-reconciler' in that an adequate account of reasoning would have to combine elements of both mental models and mental-logic accounts." Other commentators have similarly suggested a union of both models and rules (e.g., Falmagne 1993).

But, as we will show, there is no need to invoke formal rules in order to capture the necessity of valid conclusions.

The model theory provides a direct account of modal reasoning. If a conclusion holds in all of the models of the premises, then it is necessary (a valid inference from the premises); if it holds in most of the models of the premises, then it is probable (given the truth of the premises); if it holds in at least some model of the premises, then it is possible (i.e., consistent with the premises); and if it holds in none of the models of the premises, then it is impossible (i.e., inconsistent with the premises).

The model theory instantiates Smith's account of necessity. He writes: "If... your cognitive system embodies sixteen possibilities and you have eliminated all but one of them, then that remainder has a negation which is impossible. That is, it is necessary." We quoted Sherlock Holmes to the same effect: When you have eliminated the impossible then whatever remains, however improbable, must be the case (see Deduction, p. 10). Searches for mental models are fallible, but so too are searches for formal derivations - one can equally well refer to their "inductive" character. In many domains, however, it is feasible for individuals to construct all the possible models of the premises and to know that they have done so (see Barwise 1993). As an example, consider the premises:

- There is a square or else there is a triangle, but not both.
- There is a triangle or else there is a circle, but not both.

There are sixteen possibilities based on the truth or falsity of the three independent atomic propositions concerning the square, the triangle, and the circle. But these disjunctive premises eliminate all but the following two models, which each represent a separate possibility:

\[ \square \quad \triangledown \]

The putative conclusion:

There is a square and there is a circle

is not necessarily true, but only possibly true, because it fails to hold in the second model. The further premise:

There is not a triangle

rules out the second model. The premises accordingly eliminate all but one of the sixteen possibilities. "That remainder," to quote Smith again, "has a negation that is impossible. That is, it is necessary."

In logic, validity is a semantic notion, and its necessary character is established within the branch of the subject known as "model-theoretic semantics" (see, e.g., Barwise & Etchemendy 1989), which is distinct from the branch of the subject known as "proof theory." Logicians have proved meta-logical theorems showing that it is possible to frame formal rules of inference within proof theory - at least for certain logics, such as the propositional calculus - that capture all the inferences that are valid within the model theory, that is, the formal calculus is "complete." This excursion into logic shows that formal rules add nothing to a semantic account of necessity. They may...
Continuing Commentary

allow the derivation of conclusions that are necessary (according to the model theory), but in principle they cannot specify what it means for a conclusion to be necessary. A necessary conclusion is true in all the possible states of affairs in which the premises are true, but, by definition, formal rules make no reference to truth. They might, nevertheless, play some part in the development of logical ability. This idea lies at the heart of Piaget's theorizing, and he may have been right to argue that the mind develops formal operations. Yet, as the model theory shows, there is no need to postulate such a theory to account for valid deductions, and there may be no need for it in order to explain the development of logical ability in children.

Mental models and formal rules diverge in at least two other ways: on the nature of erroneous conclusions and on the possibility of a decision procedure. The failure to consider all possible models of premises leads to a characteristic error: conclusions that are possibly true rather than necessarily true. The failure to search properly for a formal derivation appears to have no systematic consequences; and it is a major weakness of formal rule theories that they have so little to say about the nature of erroneous conclusions. Reasoning based on mental models has a simple decision procedure in certain domains: a conclusion is valid if and only if it is true in all models of the premises. Reasoning based on formal rules has no decision procedure. It provides no general way, as Quine (1974, p. 75) has pointed out, to reach a verdict of invalidity (see also Barwise 1993).

R2. The search for counterexamples

Martín-Cordero & González-Labra argue that a major flaw in the model theory is its failure to specify how the mind carries out the search for counterexamples. In some domains of reasoning (as shown by the disjunctive example above), reasoners can construct all the possible models of the premises as they interpret them. In these cases and others, such as reasoning with spatial or temporal relations, the theory is in need of a search process. In reasoning that hinges on quantifiers, however, individuals are not usually able to construct all the possible models of premises as they proceed through them. Quantified assertions do not wear their logical possibilities on their sleeves. For example, the assertion:

All of the squares are not black

is ambiguous with respect to the scope of the negation. It can be paraphrased either as:

None of the squares is black

or as:

Not all of the squares are black.

Moreover, the latter interpretation is itself referentially indeterminate, that is, it is satisfied by several distinct situations even if the number of squares is held constant, such as the case where all but one of the squares is black, and, strictly speaking, the case where none of them is black. The moral is that reasoners are likely to begin by considering a proper subset of the possible models of the premises; typically, just a single model. Given the premises:

All of the Frenchmen in this restaurant are gourmets.
Some of the gourmets in this restaurant are wine drinkers.

They are likely to begin with a model that supports the conclusion:

Some of the Frenchmen are wine drinkers.

The search for counterexamples is then unlikely to be pursued assiduously, because this conclusion is so plausible. Where the same form of inference has a content that leads to the implausible conclusion:

Some of the Frenchmen are Italians.

reasoners are much more likely to discover a model of the premises that refutes it (see Oakhill et al. 1989).

Is our failure to specify how the search is carried out a major flaw in the theory? Perhaps not. The difficulty is not a conceptual one. Our computer program modeling syllogistic reasoning contains a well-specified search procedure. The difficulty is rather to obtain relevant empirical results. What the evidence shows is that individuals often overlook possible models of premises. This failure, in fact, provides an alternative account of the so-called "atmospheric effect" in syllogistic reasoning. But, so far, no evidence has revealed anything about the search process itself. The theory accordingly refrains from specifying the process. To refrain from speculation seemed like prudence rather than a major flaw.

R3. The representation of knowledge

Martín-Cordero & González-Labra argue that the criteria for everyday reasoning cannot be those of formal logic, but rather are established by semantic and pragmatic constraints. We agree: inferences in daily life do not lightly throw semantic information away by adding disjunctive alternatives to those supported by the premises. Martín-Cordero & González-Labra go on to claim that what would be considered an error from a logical standpoint might be reasonable for a lay reasoner, that such "errors" are clues to the underlying organization of knowledge, and that the model theory errs by containing no proposals about the representation of knowledge. Our view is that the strength of an everyday inference is accounted for by the model theory: given the truth of the premises, it depends in principle on the proportion of models satisfying the premises in which the conclusion is true. Where errors arise as a result of general knowledge (as in the inference above about the Frenchmen), they provide a clue to the content of general knowledge (Frenchmen are likely to drink wine) rather than to its format. The results tell us neither how the mind represents knowledge nor how it triggers pertinent knowledge during comprehension and reasoning. In fact, we have suggested that general knowledge may itself be represented in the form of mental models (Johnson-Laird 1983, pp. 68 f.). The explanatory power of the model theory may be enhanced by such a proposal, but in practice we doubt whether such conjectures have immediately testable consequences. Consider, for instance, the familiar distinction between declarative and procedural representations of knowledge. Which format is more plausible psychologically? We know of no decisive test.

The empirical phenomena of reasoning do not appear
to require explanation in terms of how knowledge is represented in the mind. Martín-Cordero & González-Labra, however, argue to the contrary: that representation is relevant to the meaning of conditionals. The model theory assumes that an indicative conditional has the same truth conditions as a material implication or equivalence, but that individuals tend to represent explicitly only the model in which the antecedent and consequent are true. Thus, inferences in the form of modus ponens (if \( p \) then \( q \); \( p \) therefore, \( q \) are easier than those in the form of modus tollens (if \( p \) then \( q \); not \( q \); therefore, not \( p \)). Because truth tables make no such prediction, it follows that the theory is not merely a notational variant of truth tables (pace Martín-Cordero & González-Labra). These authors seem about to take issue with our account of the meaning of conditionals, but their challenge veers off in another direction: “why are counterexamples more readily accepted in some content domains than others?” This issue is puzzling, but it has nothing to do with the meaning of conditionals. It concerns the fleshing out of models explicitly, perhaps as a result of knowledge—a process that has no effect on meaning, but merely on the degree to which it is represented explicitly. Again, it is an open question whether the accessibility of counterexamples has anything to do with the format of knowledge as opposed to its content.

Matters are different in the ease of knowledge of language, and the model theory does indeed include accounts of lexical semantics, parsing, and the compositional principles that underlie the representations of the meanings of sentences (see, e.g., Johnson-Laird & Byrne 1991, Ch. 9). These principles, together with other components of the theory, suffice to make three principal predictions: (1) erroneous conclusions will tend to be consistent with premises; (2) more models mean more work—reasoning will take longer and be more prone to error; (3) knowledge can affect the process of reasoning—individuals will search more assiduously for counterexamples to unbelievable conclusions. These predictions can be made a priori, and, as Deduction reports, they have been corroborated by experimental results. We therefore do not accept Martín-Cordero & González-Labra’s claim: “If a psychological theory of deduction is based on semantic procedures and does not include an explicit proposal for knowledge organization, it is unable to explain or predict reasoning performance, except in a post hoc fashion.”

R4. Conclusions

Unlike Smith, we see no reason as yet to introduce formal rules of inference into the apparatus of mental models. The theory is able to account for the necessity of valid conclusions, and for the derivation of other sorts of conclusions, such as possible and probable conclusions, which appear to be beyond the scope of current psychological theories based on formal rules. Unlike Martín-Cordero & González-Labra, we see no immediate need to make strong claims about how the mind searches for counterexamples or about how it represents knowledge. It is possible to theorize in an empirical vacuum, but there is no pressure to do so. Finally, Martín-Cordero & González-Labra refer to “the eternal debate about mental logic.” Perhaps we have concentrated too much on arguing about the virtues of mental models and the vices of formal rules. That the debate is active, we accept (see, e.g., Rips; 1994; Bonatti; 1994; in press; O’Brien et al., in press; and Johnson-Laird et al., in press; for the latest salvos). That it is useful, we wonder. That it is eternal, we doubt.

References

[1038] [1039] [1040]
[1041] [1042]
[1038] [1039] [1040]
[1038] [1039] [1040]
[1038] [1039] [1040]