approach, is incorporated in a number of texts as the systematic formal search for a counterexample as a system of formal rules. As we would all expect, the difficulty of the search increases when disjunctions of possible models are involved, and as I and most logicians would expect, the difficulty also increases when existential quantifiers are present. These particular sets of rules were chosen as much for formal goals as for psychological ease, so I am not arguing that they are exactly the right representation of ordinary deductions, but I do believe that J-L & B have not presented a conclusive case against formal rule systems generally.

The second argument is the alleged suppressibility of *modus ponens*. Given a premise "If she meets her friend, Mary will go to the play" and "Mary meets her friend" the consequent will be deduced by *modus ponens* by most deducers. However, J-L & B have found that if they also present a second premise, "If Mary has enough money, she will go to the play," that reasoners will not draw the conclusion that Mary goes to the play given that she meets her friend. The authors conclude that *modus ponens* has been "suppressed" and thus is not a mental rule. Perhaps, given the second premise, subjects mentally rewrite the first premise as "If Mary meets her friend and she has enough money, she will go to the play," in which case *modus ponens* is not suppressed but is inapplicable.

The third argument is related to the issue concerning existential quantification. J-L & B's argument that existential quantification is no harder than universal seems to have two bases – one a conceptual analysis and the other an experimental one. On page 136 they give an example of a derivation using universal quantifiers, noting that a comparable problem with existentials "differs only in that the existential quantifier, 'some', in the second premise has to be existentially instantiated, and so the quantifier restored at the end of the derivation is also existential. There is no principled way in which the derivations for the two sorts of problems can be made to differ in length."

There is no recognition that existential instantiation in many systems requires a new subproof, and that in others it requires flagging a variable or in other ways giving special status to the formula in question. (In fact, in the universal derivation there is no mention of the necessary restriction on universal generalization.) This raises doubts in my mind whether J-L & B have a sufficient grasp of what is involved in formal existential inferences.

J-L & B's experimental evidence involves two pairs of sentences. The first sentence of each is "None of the painters is related to any of the musicians," while the second sentences are, respectively:

> Some of the musicians are related to all of the authors.
> All of the musicians are related to some of the authors.

The authors report that subjects drew only 23% correct conclusions from the second pair but 64% from the first pair. They apparently conclude "Hence, there is no intrinsic difference in difficulty between existential and universal quantifiers" (p. 142), but I think that they mean to argue that the difference in difficulty cannot be explained by a difference between universal and existential quantifiers because each problem contains the same number of each quantifier. This is true, but it overlooks the fact that some proofs are much more difficult than others because of the ways in which the quantifier rules interact. In some cases the restrictions can prove a major obstacle to unsophisticated reasoners.

In any event, it is impossible to tell from their description what is transpiring because we are not told in the case where only 23% correct conclusions were reached whether the other subjects mistakenly thought no inference could be drawn or if they drew incorrect inferences. My own bet would be on the latter, since most untrained subjects have no intuitive grasp of the restrictions on quantificational inferences. Indeed, most trained subjects lose their grasp fairly quickly if they do not rehearse, and there have even been logic texts which got the subtleties wrong!

Mental models: Rationality, representation and process

D. W. Green

Department of Psychology, Centre for Cognitive Science, University College London, London WC1E 6BT, England
Electronic mail: *d.w.green@ucl.ac.uk*

It is a pleasure to read Johnson-Laird & Byrne's (J-L & B's) *Deduction*. It marshals the arguments and evidence for a mental-model theory of deduction with sustained clarity, force, and wit.

Hybrid rationality? Like theories based on mental rules, the theory of mental models proposes that there is a general competence to be explained. The arguments and experimental evidence favour the mental-model account of this general competence over a rule-based one. But is model construction and manipulation necessary for correct deductive performance? The short answer is "no." Trivially, if the answer to a problem is known, it can be retrieved. More pertinently, as J-L & B acknowledge, some individuals, tutored in logic or argumentation, may use rules or "tricks" for certain tasks. Different forms of reasoning may therefore coexist within the same individual. In addition, individuals may find shortcuts to solve specific kinds of problem. What was derived initially by envisaging a model might, during the course of the experiment, result in procedures which derive answers directly from the linguistic content. Hence, there are a variety of circumstances where model construction need not mediate rational response. If this conclusion is granted, it points to the need to consider individual patterns of performance.

Despite individual differences, I imagine that J-L & B would wish to claim that human rationality is fundamentally based on a unitary underlying competence and is not hybrid in the sense of involving both general procedures (e.g., those proposed in the theory of mental models) and domain-specific procedures, such as pragmatic reasoning schemas (Cheng & Holyoak 1985) or the cheater-detector algorithm in the social contract theory of Cosmides (1989; see also Gigerenzer & Hug 1992). Mental-model theory is, of course, more general than any domain-specific theory and is more parsimonious than any hybrid account; but neither of these properties precludes the psychological possibility that specific procedures are invoked in particular domains. It is not sufficient to show that certain findings claimed as support for domain-specific procedures can be explained *post hoc* by the theory. From an experimental point of view, more refined performance measures are required to contrast the predictions of model theory with those of domain-specific accounts of domain-specific problems, that is, of problems for which the theory of narrower scope is suited. Alternatively, empirical work on mental models could be extended to include neuropsychological data (e.g., studies of individuals with damage to the frontal lobes) that might reveal any functional dissociations (see Shallice, 1988; see also multiple book review, *BBS* 14(3) 1991), and thereby enrich the debate on the nature of the underlying cognitive architecture mediating reasoning performance. The work of Leslie and others on autism (e.g., Leslie & Thaiss 1992) confirms the possibility of dissociations in central processes.

Representational form. The procedures of model theory can be viewed as basic cognitive operations that allow the construction of models in a variety of representational forms (e.g., visuospatial). Although J-L & B rightly focus on the structural characteristics of models, it is natural to wonder about the form

in which models are represented mentally and indeed, such representations need to be specified if complete computational descriptions are to be given of performance on specific tasks. We can gain some clues by looking more closely at the process of model construction. This process treats the propositions expressed as data and constructs a mental world in which these propositions are true. Understanding is tied, temporarily at least, to the acceptance of the truth of a proposition in a way compatible with Spinoza's conjectures (see Gilbert 1991) and consistent with the way that perceptual input guides action. If there is a close relationship between thought and perception, one might expect to find correlations between performance in a perceptual domain and in a reasoning domain. Yet, as far as I know, such correlations are not obtained. Once again, neuropsychological data might prove informative. For example, subjects with deficits in visuospatial processing should perform more poorly on problems involving spatial descriptions but should not necessarily fail on syllogisms that do not reference a spatial dimension.

Processing the model. A robust finding is that performance is worse on problems that require subjects to consider more than one model. By itself, such a finding is open to two interpretations. Subjects may stop reasoning when they reach a conclusion or they may seek to envisage alternative models and fail, perhaps because of working memory constraints. In some studies, the former interpretation seems to be correct (Lee & Oakhill 1984), whereas in others (e.g., Johnson-Laird & Bara 1984), the latter interpretation seems to be correct. A crucial question, as J-L & B recognize, concerns what factors cue subjects to construct alternative models or to flesh out their initial model. They identify a number of cues, namely: The meaning of the premises may permit different initial models; initial conclusions may be considered unbelievable; the tokens depicting particular entities may be represented as not exhausting the set of such individuals. In addition, I imagine that some subjects invoke a heuristic: "Search for counterexamples." Given the variety of possible cues, it seems unlikely that there is a single psychological algorithm for evaluating conclusions. In this view, the proposed algorithm (p. 182), which first negates the conclusion and then sees whether there is an alternative model of the premises consistent with it, is one of a number.

Given the above, it seems desirable to obtain more direct evidence about the process of fleshing out the model in specific tasks so as to develop more complete accounts. In fact, a recent study which required individuals to externalize their thinking under different constraints (Green 1992) confirms the core of the mental-model account of performance of the selection task. It has also revealed an apparent paradox. Some individuals envisaged the critical counterexample but failed to select it. Such a finding implicates a postdeductive process which evaluates possible selections.

The logical content of theories of deduction

Wilfrid Hodges

School of Mathematical Sciences, Queen Mary and Westfield College, University of London, London E1 4NS, England
Electronic mail: *w.hodges@qmw.ac.uk*

Johnson-Laird & Byrne's (J-L & B's) book argues that we make deductions not by applying rules of inference to representations of the logical forms of our premises but by a process which involves building mental models of the premises and searching among them for counterexamples to the conclusion. Experiments are reported which (it is claimed) support this theory.

Let it be said at once that the mental-model theory of deduction has a pictorial quality which many people have found appealing and inspiring. Nevertheless, J-L & B's book falls short of the standards one would expect on logical writing today. There is a fair amount of symbolism, suggesting precision, but most of it is so poorly explained, or so loosely attached to the matter in hand, that the reader can only guess what is meant; time after time it happens that an interpretation which works on page *X* won't work on page *Y*.

From dozens of examples I choose two which are central. The first is an explanation of how we carry out *modus ponens;* that is, given "If p then q" and "p," how we deduce "q" (p. 47, repeated on p. 196). It is claimed that we start with the first premise, forming two mental models; the first model represents the case that p and q hold, and the second "has no explicit content." The second premise then eliminates the second model, since it is true already in the first model. Finally, from the first model we read off q. It is hard to believe that this protocol has any logical connection with the deduction that it is supposed to perform.

The second example is the notation "$[[a]b]c$" which appears on page 121 in the treatment of syllogisms. It is said to signify "that a is exhausted with respect to b, and b is exhausted with respect to c." The notion of being "exhausted with respect to something" is not explained in the text and it means nothing in logic; I dare wager it means nothing in psychology either. The interpretation which comes first to mind is that the notation means "All a's are b's and all b's are c's"; but unfortunately this reading implies that in order to use the model, we already have to be able to carry out exactly the deduction which the model was intended to explain.

This makes it impossible to comment in detail on the theory proposed in the book; I simply do not know what that theory is. Two points of methodology call for some remarks, however.

The first is the way in which J-L & B pose the basic contrast between the formal rules theory and their own mental-model theory. Supposedly these are two theories about how our minds work. But the authors tend to explain the difference by using notions from the mathematical theory of formal systems. A typical example is on page 212, where they explain that mental models "do not contain variables." Without some explanation of what it is for a mental representation to "contain a variable," this is meaningless. (My own impression is that many of the mental models described in this book do in fact contain components which behave pretty much like variables, if one looks at variables in the appropriate way.) Because of this mismatch between the phenomena to be explained and the concepts used to explain them, the book fails to establish a genuine difference between formal rules and mental models.

The second point of methodology concerns the claim that a theory of deduction based on mental models "predicts which problems will be difficult and it predicts which errors ordinary individuals will make with them" (p. 131). This claim will not survive a closer look at what is meant by "a theory based on mental models." Take, for example, the case of syllogisms, as in Chapter 6. If the theory in question is either (1) the general theory that we make deductions by forming models of the premises and looking for counterexamples to the conclusion, and so on, or (2) the theory of models of syllogisms as presented in the chapter, then it is too imprecise to have the consequences claimed, for example about the numbers of models needed for each syllogism.

One suspects that the authors may have in mind (3), the detailed theory propounded in Johnson-Laird & Bara (1984). This theory is different from the one outlined in the chapter, but it seems to underlie some of the discussion, and it is precise enough to be written as a computer program. The problem with this third theory is that it involves, among other things, fourteen "principles" for carrying out operations, some of them more *ad hoc* than others. Since the theory has almost as many degrees of freedom as the data to be explained, the reasonable fit is hardly impressive. To justify their claim, the authors need to produce a theory which is precise enough so that the reader can verify what predictions it makes, and one that is also derivable from