

Mental models and nonmonotonic reasoning

Nick Chater

Department of Psychology, University of Edinburgh, Edinburgh EH8 9JZ, Scotland

Electronic mail: *nicholas@cogsci.ed.ac.uk*

Johnson-Laird & Byrne (J-L & B) are equivocal concerning the scope of mental-model theory. On the one hand, they are careful to note that mental models are aimed primarily at explaining deduction, although commonsense inference is not deductive in character. On the other hand, they contend that mental-model theory solves the problem of nonmonotonic reasoning, which is *not* deductive and is characteristic of commonsense inference. This equivocation requires clarification: An account of deductive reasoning casts light on a fascinating if rather arcane human ability; an account of nonmonotonic inference in general would be little short of a theory of thinking. It is not clear, therefore, exactly what J-L & B see as the domain of the mental-model account. I shall argue that mental-model theory does not in fact address the problem of understanding commonsense nonmonotonic reasoning, still less provide a solution to it.

Everyday, commonsense reasoning may be conceived of as a species of inference to the best explanation: It involves inferring from given information to what best explains and is best explained by that information (Fodor 1983). Such inference is nonmonotonic, because the addition of new information can invalidate what were previously plausible conclusions. So, for example, the plausible inference from hearing the sound of purring behind the door to the conclusion that the cat is trapped in the cellar is immediately overridden if I catch sight of the cat in the garden. The premise on which the inference is based, the purring, need not be withdrawn, although another explanation for this fact may be sought. By contrast, in monotonic reasoning, the conclusion of a valid argument can only be challenged if one of its premises is false.

Providing an account of inference to the best explanation is very difficult. Inference to the best explanation encompasses

theory confirmation in science, the problem of inferring the scientific theory which best fits the available data. Understanding inference to the best explanation also presupposes a solution to the notorious frame problem (McCarthy & Hayes 1969). Indeed the problem of nonmonotonic reasoning and the frame problem really constitute the same problem looked at from different points of view. Where nonmonotonic reasoning concerns which conclusions *should* be revised when new information is added, the frame problem involves deciding which conclusions *need not* be revised.

Nonmonotonic reasoning has proved resistant to a vast number of extremely ingenious proposals within artificial intelligence. The principal line of attack has been to attempt to provide a *nonmonotonic* logic which captures commonsense, rather than deductive, inferences (e.g., Hanks & McDermott 1987; McCarthy 1980; Reiter 1980; Shoham 1987). These logics have been dogged by two serious problems (McDermott 1987; Oaksford & Chater 1991). First, it has not been possible to capture the nonmonotonic inferences that people routinely draw (specifically, most computational methods tend only to draw extremely weak conclusions). Second, these methods have been plagued by computational intractability due to explicit or implicit reliance on consistency checking. J-L & B's treatment of nonmonotonic reasoning addresses neither of these difficulties, but I shall concentrate on the first and most fundamental.

Throughout their book J-L & B give examples of how mental models can be constructed and checked to generate inferences licensed by a standard logic (typically, propositional or predicate calculus). This is not surprising, because they are primarily concerned with deductive reasoning. The rules governing the building of models and searching for alternative models ensure that the notion of the validity of the logic concerned is respected. A mental-model account of nonmonotonic reasoning would require similar rules for model construction and search; but how could these rules be chosen? By analogy with the deductive case, we might expect these constraints to respect the notion of the validity of some nonmonotonic logic. J-L & B rightly do not attempt to follow this line, since there is no appropriate nonmonotonic logic to which appeal can be made. Yet, without an underlying logic of some kind, the processes of building mental models will be without any constraint or justification.

There is also a more specific worry. Inference on the basis of a failure to find countermodels, which is at the heart of the mental-model account, appears inapplicable to nonmonotonic reasoning, because, by definition, such countermodels invariably exist: If there were no model in which the premises are true and the conclusion false, the reasoning would be valid according to a standard, deductive, monotonic logic, and would not be an instance of nonmonotonic reasoning at all. Furthermore, not only do countermodels exist, but people find it easy to generate these countermodels on demand. For example, returning to the inference that the cat is in the cellar on the basis of hearing a purring sound, it is easy to generate a host of countermodels in which this inference is not correct. For example, the sound may be produced by a different cat, a tape recorder, a person doing a cat imitation, the wind, and so on. Quite generally, it is easy to find countermodels for everyday nonmonotonic inferences. This means that if nonmonotonic inference proceeded from a failed search for countermodels, no nonmonotonic inferences would be drawn at all! Notice that the mental-model theorist cannot argue that only plausible models are constructed, or that the single most plausible model is chosen, because the problem of finding such models is simply a restatement of the problem of inference to the best (i.e., most plausible) explanation (i.e., model). If mechanisms for assessing plausibility could be assumed, then the problem of nonmonotonic reasoning would already have been solved, and invoking mental models would be unnecessary (for a discussion of more detailed proposals concerning how mental models might be applied to nonmonotonic reasoning, see Chater & Oaksford 1993; Garnham 1993).

What leads J-L & B to claim that they have solved the problem of nonmonotonic reasoning? Simply the observation that inferences made on the basis of a particular model may not follow deductively from the premises given; the addition of later information which demands that the set of models be searched more fully can lead to such a conclusion's being withdrawn. This certainly shows how mental models can make logically invalid inferences and later recover from them, but it says nothing about the real problem in hand. The problem of nonmonotonic reasoning is not to explain how it is possible to jump to conclusions and later revise them; this could be achieved trivially by modifying any standard method of proof. The problem is to explain how people are able to jump to *sensible* conclusions and revise these *appropriately* as new information is encountered, and indeed to elucidate what it means for a conclusion to be sensible or appropriate. Regarding these questions, artificial intelligence and cognitive science have had depressingly little to say; unfortunately, mental-model theory appears to have nothing further to add.