

How We Reason: A View from Psychology

Psychologists have studied reasoning for at least a century. But, for sixty years or so, they had no proper theory of what individuals are doing when they reason, or of the underlying mental processes, which are inaccessible to introspection. Computers made theorizing about reasoning feasible and respectable, and psychologists have developed several such theories, especially of deduction. One theory is that we are all equipped with formal rules of inference akin to a logic in a “natural deduction” formulation, see, e.g., Rips, L. (1994: *The Psychology of Proof*, Cambridge, MA: MIT Press). Reasoning on this account is a search for a proof leading from premises to conclusion. Another theory is that the probability calculus describes how we ought to reason and how in fact we do reason even deductively, see Oaksford, M., and Chater, N. (2001: The probabilistic approach to human reasoning, *Trends in Cognitive Sciences*, 5, 349–357). This theory describes well the results of certain psychological experiments, yet other experiments have shown that untrained individuals do distinguish between necessary conclusions and probable conclusions. Indeed, Louis Lee and Geoffrey Goodwin have shown in unpublished studies of Sudoku puzzles that naïve individuals soon realize that their solution depends, not on probabilities, but on valid deductions.

One difficulty for theories based on formal rules is that reasoning in daily life depends on the logical form, not of sentences in natural language, but of the propositions that they express. Hence, the use of formal rules depends on recovering the logical form of propositions.

For example, in the sentential calculus, an inference of this form is valid:

- If p then not q .
- q .
- Therefore, not p .

But, not surprisingly, individuals balk at this inference:

- If Jane played a game then she didn't play soccer.
- Jane did play soccer.
- Therefore, she didn't play a game.

They know that soccer is a game, and therefore that the conditional premise is consistent with only two possibilities, shown here on separate lines:

- Jane played a game. Jane didn't play soccer.
- Jane didn't play a game. Jane didn't play soccer.

The logical form of the first premise is therefore: (p or not p) and not q . Its recovery is a headache, and the general analysis of the logical form of propositions expressed in natural language is beyond any existing algorithm.

If my colleagues and I are correct, there is no need to recover logical forms and no need to search for proofs. Our theory postulates instead that individuals use the meanings of propositions and general knowledge to construct a set of mental models representing the possibilities consistent with the premises, see Johnson-Laird, P.N., and Byrne, R.M.J. (1991: *Deduction*, Hillsdale, NJ: Erlbaum) and Johnson-Laird, P.N. (2006: *How We Reason*, Oxford: Oxford University Press). If a conclusion holds in all these possibilities, then individuals infer that it is valid. And they are also able to show that an inference is invalid, not by searching in vain for its derivation, but by constructing a counterexample, i.e., a model of a possibility consistent with the premises but not with the conclusion.

One prediction of our theory is that the greater the number of models needed to make an inference, the harder it will be. For example, ask yourself what follows from these two disjunctions:

- Ann is in Atlanta or Ben is in Birmingham, or both.
- Ben is in Birmingham or Cate is in Clemington, or both.

Intelligent but untutored individuals usually overlook at least one of the five possibilities consistent with these premises, and their conclusions often describe only one of them.

A fundamental principle of the theory is that mental models represent what is true, but not what is false. Hence, given the first premise above, individuals enumerate the following three possibilities:

- 1. Ann is in Atlanta.
- 2. Ben is in Birmingham.
- 3. Ann is in Atlanta. Ben is in Birmingham.

where, for example, the falsity of first disjunct in the second possibility is not represented explicitly, see, e.g., Johnson-Laird, P.N., and Savary, F. (1999: Illusory inferences: A novel class of erroneous deductions, *Cognition*, 71, 191–229). This so-called principle of *truth* reduces the processing load on working memory, but, as we discovered from a computer program implementing the theory, it has a devastating effect on certain seemingly simple inferences. Consider this problem, for instance:

- Either Jane is kneeling by the fire and she is looking at the TV, or else Mark is standing at the window and he is peering into the garden.
- Jane is kneeling by the fire.

Does it follow that she is looking at the TV?

Most individuals say, “yes”, see Walsh, C., and Johnson-Laird, P.N. (2004: Co-reference and reasoning. *Memory & Cognition*, 32, 96–106). Given the first premise, they think of two possibilities: in one, the first conjunction is true; and in the other, the second conjunction is true. They overlook that when the second conjunction is true, the first conjunction is false, and that one way in which it can be false is when only its first clause is true, i.e., Jane is kneeling by the fire but *not* looking at the TV. Hence, the correct answer to the question is: “no”.

Invalid inferences of this sort are endemic, occurring in all domains of reasoning. Untutored individuals represent what is true (rather than what is false), what is possible (rather than impossible), what is permissible (rather than impermissible, except in the case of overt prohibitions), and what are instances of a concept (rather than non-instances). It is as though for them what is false etc. ceases to exist. Any theories, including those based on formal rules or on the probability calculus, that fail to predict this phenomenon have quite a bit of explanatory work to do. There is, of course, more to the model theory than I can describe in this outline—it extends to probabilistic reasoning, induction, and abduction. Likewise I am indebted to more colleagues than I can name here.

P.N. Johnson-Laird
Psychology, Princeton