The Search for Counterexamples in Human Reasoning

Hansjörg Neth (hneth@princeton.edu)
Philip N. Johnson-Laird (phil@clarity.princeton.edu)
Department of Psychology; Princeton University
Princeton, NJ 08544, USA

A major point of contention about human reasoning is whether or not individuals search for counterexamples to conclusions. According to theorists who argue that the mind is equipped with tacit rules of inference, the decision that an argument is invalid depends on a failure to find a derivation leading from the premises to the conclusion (see e.g. Rips, 1994). However, with this procedure, one can never be certain that the space of possible derivations has been searched exhaustively.

Alternatively, reasoners may base their inferences on mental models (Johnson-Laird and Byrne, 1991). This theoretical account rests on the semantic principle of validity: a conclusion is valid if and only if it allows for no counterexamples, i.e., possibilities in which the premises are true but the conclusion is false. Hence, by constructing a counterexample, reasoners are able to *know* that an inference is invalid.

So, how do logically naïve individuals establish invalidity? Surprisingly for so central a question, there is a dearth of evidence. In the case of syllogisms, Polk and Newell (1995) have defended an account in terms of models, but argued that their explanation of individual differences gains little, if anything, by postulating a search for counterexamples. Bucciarelli and Johnson-Laird (1999) have shown that people are able to search for counterexamples if prompted to do so. Whether individuals spontaneously engage in this search is still unknown.

An experiment on reasoning with non-standard quantifiers

To search for a search for counterexamples, we carried out an experiment in which the participants had to evaluate eight inferences based on *non-standard* quantifiers. Quite common in everyday life, such inferences call for a higher-order predicate calculus, which is incomplete. Each problem had one, two, or three premises, each based on the quantifier "more than half", and a putative conclusion based on "at least one" or "more than half", e.g.:

> More than half of the visitors speak English.
> More than half of the visitors speak French.
> More than half of the visitors speak German.
> Therefore, at least one visitor is trilingual.

Because our main interest was the search for counterexamples, five of the eight inferences were invalid. 20 Princeton undergraduates were instructed to think aloud as they tackled the inferences in a random order. Each problem was presented on a separate sheet of paper on which the participants were encouraged to write or draw whatever would help them.

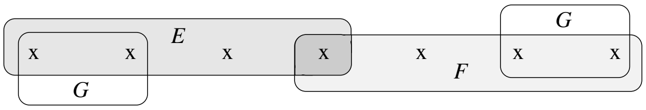

Figure 1: A visual counterexample in which each x represents an individual, more than half of them have features *E*, *F*, and *G*, but it is false that at least one x has all three features.

Overall, participants responded correctly in 72% of all trials. The two-premise problems were easiest (95% correct), the one-premise problems were intermediate in difficulty (73%), and the three premise problems were hardest (48%), and this trend was reliable (Kendall's W=.56, p<.01).

A striking phenomenon was the participants' spontaneous use of a variety of strategies. In 80% of all trials, they constructed a single specific instance of the premises. Typically, the participants constructed a diagram in which they tried to *minimize* the overlap between the given sets. Indeed, every single participant came up with at least one counterexample (see Figure 1). The participants evaluated inferences as invalid on 69 trials and were correct on 62 of them. They produced a counterexample (58% of these trials), claimed that a counterexample was possible (10%), or used some method that the protocols did not reveal (32%). Likewise, they frequently used an unsuccessful search for counterexamples as a basis for concluding that an inference was valid. Where the protocols revealed the nature of the participants' strategies, they relied on counterexamples more often than not (60% of all trials, Wilcoxon test, N=13, T^+=83.5, p<.003).

We still know very little about what governs the selection of strategies. In general, one domain of reasoning may elicit a search for counterexamples more readily than another. Also, the evaluation of a given conclusion may be more likely to trigger a search than the formulation of one's own conclusions. Our study has shown, however, that logically naïve individuals do spontaneously search for counterexamples for at least one sort of deduction.

References

Bucciarelli, M., & Johnson-Laird, P. N. (in press). Strategies in syllogistic reasoning. *Cognitive Science*.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction.* Hove, England: Lawrence Erlbaum Associates.

Polk, T. A., & Newell, A. (1995). Deduction as verbal reasoning. *Psychological Review*, *102*, 533–566.

Rips, L. J. (1994). *The psychology of proof: Deductive reasoning in human thinking.* Cambridge, MA: MIT Press.