# Bounded Rationality in Taking Risks and Drawing Inferences

**Mike Oaksford**
UNIVERSITY OF WALES, BANGOR

**Nick Chater**
UNIVERSITY OF EDINBURGH

ABSTRACT. This commentary provides a discussion of the concept of 'bounded rationality' as it applies to the theses advanced by Lopes (1991) and Evans (1991). Lopes's (1991) assessment of the irrationalist consequences of Tversky and Kahneman's (1974) work on heuristics and biases is premature because bounded rationality implies that people *could not* employ optimal strategies. Considerations of bounded rationality also provide additional criteria by which to judge the theories of deductive reasoning discussed by Evans (1991). Judged by this criterion, theories whose goal is to explain logically competent performance are inadequate (Oaksford & Chater, 1991). Thus Evans's assessment of the state of current theories of reasoning requires revision.

This commentary is on two separate articles which appeared in *Theory & Psychology*, volume 1(1), by Lola Lopes (1991) and Jonathan Evans (1991). Our reasons for offering a joint commentary is that in both papers an issue appears to be overlooked which has potentially serious consequences for the theses each author was concerned to advance. We begin with the article by Lopes.

## Heuristics and Biases

Lopes (1991) criticizes work in the 'heuristics and biases' tradition because the rhetorical emphasis of the papers reporting this work has led to an overestimation of human irrationality. The original papers (Kahneman & Tversky, 1972, 1973; Tversky & Kahneman, 1971, 1973), Lopes argues, were about the *processes* involved in spontaneous judgements in risky decision-making: were suboptimal heuristics being employed or were optimal algorithmic procedures being used? Lopes observes that in the summary article on this early work by Tversky and Kahneman (1974) in *Science*, the emphasis changes from process to cognitive *bias*. Rather than discuss the successes of the quick and dirty heuristics they discovered, Tversky and Kahneman (1974) dealt at length with the lapses from optimal rationality to which the use of such heuristics may lead. As Lopes observes, this

THEORY & PSYCHOLOGY © 1992 SAGE. VOL. 2(2): 225–230

emphasis set the tone for much subsequent discussion, leading to possibly premature conclusions about the irrationality of human decision-making and reasoning processes.

Interpreting the influence of a body of work may often depend upon the perspective adopted. From the perspective of computational modelling there is an interpretation of the heuristics and biases literature which fails to lead to any particularly dire conclusions for human rationality. Kahneman and Tversky were working within the framework of 'bounded rationality' which they attribute to Jerome Bruner and Herb Simon (see the Preface to Kahneman, Slovic, & Tversky, 1982). The nature of these bounds can best be understood by taking into account the constraints placed on cognitive processes by the claim that they are computational processes. A major constraint is that these processes must be capable of utilization within the time-scale at which normal human judgements are made. In computer science these issues are discussed under the heading of *computational complexity theory* (see, e.g., Garey & Johnson, 1979). Some computational processes are more complex than others, requiring more computational resources in terms of memory capacity and operations performed. Measures of complexity are expressed as a mathematical function relating the length of an input ($n$)—very roughly the amount of information which the process must take into account—and the amount of computational resources consumed. Any process which requires exponentially increasing resources (i.e. increasing at a rate of $2^n$, or worse) are regarded as computationally intractable. That is, for some $n$ these processes may not provide an answer in our lifetimes if at all.

Issues of computational complexity have cropped up quite frequently in the history of cognitive psychology and artificial intelligence, perhaps most notably in vision research. Early work on bottom-up object recognition of blocks worlds resulted in the notorious combinatorial explosion (see McArthur, 1982, for a review, and Tsotsos, 1990, for a more recent discussion of complexity issues in vision research). In the research into risky decision-making, it was realized very early that complexity issues were relevant. Bayesian inference makes exponentially increasing demands on computational resources even for problems involving very moderate amounts of information. A salutary example is provided by the discussion of an application of Bayesian inference to medical diagnosis problems involving multiple symptoms in Charniak and McDermott's (1985) *Introduction to Artificial Intelligence*. Diagnoses involving just two symptoms, together with some reasonable assumptions concerning the numbers of diseases and symptoms a physician may know about, requires upwards of $10^9$ numbers to be stored in memory. Since typical diagnoses may work on upwards of 30 symptoms, even if every *connection* in the human brain were encoding a digit, its capacity would none the less be exceeded.

Spontaneous, real-world risky decisions, even of moderate complexity, are not made using Bayesian inference processes, because they *could not* be. Since the mind/brain is a limited information processor, the processes of risky decision-making cannot be based upon optimal, algorithmic procedures. This means that the only rationality to which we can aspire, as individual decision-makers, is one bounded by our limited computational resources. In consequence, the observation that we do not behave in accordance with Bayes's theorem could not impugn our rationality. Our rationality could be questioned only if we were capable of using the



optimal strategy but failed to do so. Thinking otherwise is akin to condemning us because we do not fly even though we do not possess wings.

Three further issues deserve mention. First, Lopes (like us) is concerned only with individual decision-making, without pencil, paper (computer) or friends, as it were. The additional resources available in groups and societies mean that decision-making can transcend the limitations of the individual. The existence of Bayes's theorem is a testament to the collective rationality of a culture embodied in modern mathematics. Second, it could be argued that the laboratory tasks employed by Kahneman and Tversky would have permitted the use of the normative strategy because the amount of information ($n$) was kept well within manageable bounds. Thus, the fact that the heuristics were still employed may have some negative implications for human rationality. However, with no schooling in statistics, the only strategy available is to generalize those strategies normally employed in more complex settings to the laboratory task. Restricting the information could only encourage the use of Bayes's theorem if it had been previously learned. Third, Lopes adduces evidence (Gigerenzer, Hell, & Blank, 1988) that when some problems are presented more realistically, subjects do take account of prior probabilities in accordance with Bayes's theorem. From the perspective of bounded rationality, of course, it is such apparent displays of competence which create a problem since (i) they do not cohere immediately with the heuristic approach and (ii) they *could not* be a product of a general, unlearned competence with Bayes's theorem.

In summary, considerations of bounded rationality temper the irrationalist consequences of the work on heuristics and biases. Only by ignoring bounded rationality could the rhetoric of Tversky and Kahneman (1974) be interpreted as leading to the dire conclusions drawn by Lopes in her article. Given the unjustifiable presumptions of normative rationality which were rife in the psychological literature at the time, the rhetorical bias of Tversky and Kahneman's summary article may have set just the right balance to provide a much needed corrective.

## The Fragmented State of Reasoning Theories

The deductive reasoning literature reviewed by Jonathan Evans (1991) raises directly analogous issues concerning human rationality as we have seen above in the area of decision-making under uncertainty. Evans's paper discusses the way that research into deductive reasoning has fragmented of late, with different theories answering different questions raised by the data. He observes that there are three questions which need to be answered: the competence question—the fact that human subjects often successfully solve deductive reasoning problems; the bias question—the fact that subjects also make many systematic errors; the content and context question—the fact that the content and context of a problem can radically alter subjects' responses. The major theories in this area—mental logics, mental models, schema theories and heuristic approaches—all tend to concentrate upon one question or the other, none providing a fully integrated account of all three. Evans does, however, provide criteria of theory preference—completeness, coherence, falsifiability and parsimony—by which to judge reasoning theories, and

seems to view mental models as scoring most highly on these criteria. Evans's paper is an important and laudatory attempt to get reasoning theorists to agree some common ground rules concerning the adequacy of their theories. However, an additional criterion of theory choice may place a very different complexion on the adequacy of current theoretical proposals.

Bounded rationality is not an issue which is frequently discussed in the deductive reasoning literature. However, issues of computational complexity may serve as a valuable criterion for choosing between reasoning theories in addition to the general criteria proposed by Evans which are common to all areas of scientific inquiry. To the extent that issues of resource limitation are mentioned in the reasoning literature, they are restricted to discussion of how our limited short-term memory capacity may lead to systematic errors in explicit reasoning tasks (Johnson-Laird, 1983). However, one reason why the deductive reasoning literature has been so prominent within cognitive psychology/science is the assumption that the principles of human inference discovered in the investigation of explicit inference will *generalize* to provide accounts of all inferential processes. This is important because, qua computational process, *all* cognitive processes can be viewed as inferential (Boolos & Jeffrey, 1980). We will call this the *Generalization Assumption*. The generalization assumption is, for example, embodied in the subtitle to Johnson-Laird's (1983) book *Mental Models: Towards a Cognitive Science of Language, Inference and Consciousness*. Without the generalization assumption, the study of deductive reasoning would warrant little more interest than, say, the psychology of doing crosswords.

In artificial intelligence, studying theories of inference and knowledge representation usually begins by examining their capabilities in *toy* domains. Toy domains are specially contrived micro-worlds about which very little needs to be assumed. There is, however, a long-standing problem with this approach. Theories of inference which are adequate in such domains (e.g. the inference engine in SHRDLU: Winograd, 1972) tend to fail disastrously when they are *scaled up* to deal with real-world inferential problems involving more information (higher $n$). This is because they are generally computationally intractable. A directly analogous issue arises for psychological theories of reasoning designed to account for laboratory tasks but with pretensions to satisfy the generalization assumption. *All* the theories which attempt to answer Evans's competence question hit computational intractability problems when scaled up to deal with real-world inferential problems (Oaksford & Chater, 1991). It is, moreover, a recent realization that even in explicit reasoning tasks the range of information ($n$) taken into account in drawing an inference transcends that explicitly provided in the task. As Evans observes, Johnson-Laird and Byrne's (1991) 'fleshing out' strategy involves the incorporation of more information, derived from prior world knowledge, to supplement that explicitly provided, as does the addition of implicit premises in a mental logic account. Oaksford and Chater (1991) point out that logics based on syntactic proof procedures, like those proposed in mental logic accounts, are computationally intractable in everyday inferential contexts. Moreover, semantic proof procedures, like mental models, are known to be *worse* in complexity theoretic terms than syntactic procedures. Hence the two major contenders to answer the competence question may not only fail to satisfy the generalization assumption, to the extent that explicit inference relies upon 'fleshing



out', they may also be poor contenders as theories of laboratory reasoning tasks.

In summary, a bounded rationality assumption may also need to be made in theories of deductive reasoning. On analogy with Bayes's theorem in decision-making under uncertainty, our ability to perform in accordance with logical dictates cannot be taken as evidence that we possess a general unlearned logical competence, *if*, by logical competence, we mean that we employ a logical system in our reasoning, be it syntactically *or* semantically realized. Again, in the general case, this is because we *could not* be using such a system, and again, therefore, that we occasionally deviate from logicality could not impugn our rationality. As mentioned above, this means that Evans's competence question is the problematic one. Again it is reasonable to assume that whatever quick and dirty mechanisms we have evolved in order to resolve the complex inferential problems of everyday reasoning will also be generalized to the laboratory tasks studied by reasoning theorists. However, apart from Evans's own proposals concerning the use of heuristics in the interpretation of premises, we appear to remain profoundly ignorant of the nature of these mechanisms.

**References**


Boolos, G., & Jeffrey, R. (1980). *Computability and logic* (2nd ed.). Cambridge: Cambridge University Press.

Charniak, E., & McDermott, D. (1985). *Introduction to artificial intelligence*. Reading, MA: Addison-Wesley.

Evans, J. St B.T. (1991). Theories of human reasoning: The fragmented state of the art. *Theory & Psychology, 1*(1), 83–105.

Garey, M.R., & Johnson, D.S. (1979). *Computers and intractability: A guide to the theory of NP-completeness*. San Francisco, CA: W.H. Freeman.

Gigerenzer, G., Hell, W., & Blank, H. (1988). Presentation and content: The use of base rates as a continuous variable. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 513–525.

Johnson-Laird, P.N. (1983). *Mental models: Towards a cognitive science of language, inference and consciousness*. Cambridge: Cambridge University Press.

Johnson-Laird, P.N., & Byrne, R.M.J. (1991). *Deduction*. Hillsdale, NJ: Erlbaum.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3*, 430–454.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*, 237–251.

Lopes, L.L. (1991). The rhetoric of irrationality. *Theory & Psychology, 1*(1), 65–82.

McArthur, D.J. (1982). Computer vision and perceptual psychology. *Psychological Bulletin, 92*, 283–309.

Oaksford, M., & Chater, N. (1991). Against logicist cognitive science. *Mind and Language, 6*, 1–38.

Tsotsos, J.K. (1990). Analyzing vision at the complexity level. *Behavioral and Brain Sciences, 13*, 423–469.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76*, 105–110.

Tversky, A., & Kahneman, D. (1973) Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5*, 207–232.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124–1131.
Winograd, T. (1972). *Understanding natural language*. New York: Academic Press.

MIKE OAKSFORD is Lecturer in Cognitive Science at the Department of Psychology, University of Wales, Bangor. He is also Co-Director of the Cognitive Neurocomputation Unit at the same university. ADDRESS: Cognitive Neurocomputation Unit, Department of Psychology, University of Wales, Bangor, Gwynedd LL57 2DG, UK.

NICK CHATER is Lecturer in Cognitive Psychology at the Department of Psychology, University of Edinburgh, Scotland. He is also an Associate member of the Centre for Cognitive Science at the same university.