

The Problem of Irrationality

Lorraine Daston

Department of History, Brandeis University, Waltham, MA

Gerd Gigerenzer

Department of Psychology, University of Constance, Germany

Why are people so stubbornly and extravagantly irrational, so often ignoring not only sound logic but also solid self-interest? This age-old philosopher's sigh is the departure point for Howard Margolis' stimulating, synthetic work on the psychology of thinking in general and of scientific discovery and persuasion in particular. Margolis advances a strikingly original view about how the mind works: in contrast to fashionable psychological approaches that draw analogies between thinking and computer programs, Margolis insists that all thinking and judgment can be "reduced to pattern recognition" (p. 1). Rules, algorithms, computation, and logic play no role in this "*P*-cognition," although the incessant shuttling back and forth between situational cues and our stored repertoire of patterns may ultimately produce all of these. But however logical and rational its incidental products (p. 94), the brain itself is, on Margolis' account, "a-logical and a-rational" (p. 4).

Margolis' answer to the venerable problem of irrationality is an evolutionary one: we are irrational because we are made that way. In successful species, cognition balances between the advantages of efficiency and of careful attention to particular situations. *P*-cognition, suitably disciplined by "checking," has the virtues of speed tempered by a modicum of caution. It also has the vices of those virtues: the more entrenched a pattern, the more efficiently it can be triggered in a new situation—and the more difficult it will be to correct as a truly new situation will demand. We err, be it by succumbing to cognitive illusions or through conservatism in the face of scientific novelty, because we apply the wrong patterns, misled by the ambiguity of cues, the tyranny of mental habit, or both. Thus, the same processes that are responsible for our cognitive successes—"our facility in handling patterns in faces, languages, places, and even entirely artificial things like chess and concert music" (p. 105)—are also to blame for our cognitive failures.

Margolis is largely mute on how we acquire the patterns of *P*-cognition, how we select among the many such patterns in our repertoire and what exactly pattern recognition entails. And wisely so, for his brief attempt (chapter 4) to clarify this last plunges him into contradiction. How to reconcile Margolis' analogy between *P*-cognition and a computer curve-fitting program with his earlier insistence that cognition has nothing to do with rules, algorithms, and calculations? Or with his evolutionary "cognitive ladder," which ascends from "simple feedback" through reasoning, with calculation as the most recent and most tenuously established achievement (p. 44)? The fatal attraction of the computer curve-fitting analogy for Margolis seems to be its openness to inspection and analysis, and the clarity these promise (Margolis himself is an admirably clear writer).

Margolis hopes that the evolutionary outlook of *P*-cognition will prove its worth by yielding new interpretations of old problems in the psychology of cognitive illusions and in the history of scientific revolutions. What can *P*-cognition tell us about hotly debated cognitive illusions produced, for example, by the Wason selection task and the Kahneman and Tversky cab problem? Margolis suggests it is equally wrong to conclude from these experiments that the brain is, on the one hand, incapable of heeding *modus tollens* or base rates, or, on the other, that the laws of thought are the laws of logic. Rather, subjects in these experiments react to ambiguities in language and “scenario” by applying otherwise valid patterns to the wrong situation: they give the right answers to the wrong questions. Thus far, Margolis’ interpretation persuasively challenges theorists from Piaget to Fodor to L. J. Cohen who liken thinking to logic and calculation, replacing these implicit logical rules with “habits of mind” prompted by context.

However, Margolis nonetheless insists that these context-cued judgments deserve to be called “illusions,” because misreading admittedly ambiguous cues still counts as irrationality to him. This view is in turn anchored in the conviction that, for example, the cab problem has only one “right answer” as far as statisticians are concerned, however much philosophers and psychologists who “should know better” may bicker about it (p. 306). But here it is Margolis who is deluded by his own preoccupation with cut-and-dried standards of rationality: there are in fact perfectly respectable alternative solutions to the cab problem, for example those using Neyman-Pearsonian statistical methods. Some, indeed most, situations are genuinely ambiguous. Unfortunately, this is no passing slip, for it seriously handicaps Margolis’ attempts to analyze cognitive illusions in *P*-cognitive terms, leading him to emphasize his unconvincing accounts of how scenario ambiguities produce illusions over his convincing accounts of how language ambiguities do—presumably because subjects can hardly be judged irrational for misconstruing equivocal wording. In his interpretation of the cab problem, for example, this preference for scenario ambiguities obliges Margolis to make the dubious claim that reasoning from causes to effects is more habitual than reasoning from effects to causes.

What can *P*-cognition tell us that we didn’t already know about such well-researched episodes in the history of science as the Copernican Revolution? Margolis contends that Aristotelian physics, religious qualms, common sense, and (most surprisingly) the equant played at best a secondary role in inhibiting the initial heliocentric insight and its subsequent spread. Rather, the tenacious pattern of contiguous, “nested” spheres that allowed Ptolemaic astronomers to order the planets and estimate the relative distances between them (albeit somewhat arbitrarily, especially in the case of the inferior planets) blocked both Copernicus’ Arab predecessors and his immediate successors from grasping the (for Margolis) patent advantages of coherence offered by heliocentrism. On this account, both Tycho’s shattering of the heavenly spheres and his compromise system are a grand anticlimax—far more decisive was “the mere passing of time” between 1543 and the 1580s, allowing a new generation of astronomers to “come to maturity always aware of the Copernican possibility” and to find that “the Copernican argument explains so much so neatly that it can hardly be wholly wrong” (p. 257).

Alas, there is scant evidence to support this pattern-centered account, and a considerable amount that contradicts it. On the iconographic evidence of the sort Margolis mostly invokes, the famous diagram in 1.10 of *De revolutionibus* strongly suggests that Copernicus himself still clung to the contiguous sphere scheme, although it was no longer necessary to compute relative planetary distances. Neither Copernicans nor their adversaries made much of the nested spheres issue, and Kepler displayed a near lunatic affection for cozy nested patterns in his ardently Copernican *Mysterium cosmographicum*. At a deeper level, Margolis’ reading of the Copernican revolution betrays the same crippling preoccupation with wrong and right answers that distorts his

interpretation of cognitive illusions. Although he struggles to see the past in its own terms, he cannot resist reproaching early opponents of Darwin and of Copernicus for their cognitively explicable but rationally indefensible failure to embrace the new theories—Lyell’s resistance to natural selection is “cognitively not much different from (not much more logical than)” resistance to wearing hockey helmets (p. 192), just as subjects in psychological experiments are “stubborn” and “downright stupid” in not applying *modus tollens* or Bayes’ theorem.

In both cognitive psychology and the history of science, this normative *idée fixe* prevents Margolis from posing more interesting descriptive questions about how we cope with ambiguity and how the standards for explanation, not just this or that explanation, change in science. Although an abiding concern with rationality motivated Margolis’ study, there seems to be nothing in his *P*-cognition view *per se* that would commit him to such a narrow normative stance. Margolis believes his account will prove itself by its empirical applications in fields like cognitive psychology and history of science. We believe, on the contrary, that the results of these initial empirical applications are largely disappointing, but that the general Darwinian outlook of *P*-cognition so ably developed by Margolis will prove of enduring value.