To appear in I. Weiner, & E. Craighead (Eds.) *The Corsini encyclopedia of psychology*, 4th Edition. Hoboken, NJ. Wiley & Sons.

Joachim I. Krueger
Brown University

Rationality

With the study of human cognition comes the question of how well people can reason, and any answer requires standards with which observed cognitive performance can be compared. Cognition that does not meet these *a priori* standards is considered irrational. The Socratic philosophers held that rationality prevails unless disrupted by instincts, appetites, or emotions. They believed that only rational thought, but not momentary affect, can determine the best course of action. People are being irrational when they act against their own best interest, rationally defined. This conflict theory pervades Western philosophy and psychology. Of the topics for psychological research it has inspired, delay of gratification and temporal discounting are good examples. The theory holds that the rational mind takes the long-term view of the organism's well-being, whereas irrational forces demand rapid gratification. Although discounting of the future is not by itself irrational, its steep hyperbolic nature is. It is incoherent to prefer $50 now to $100 in a year, while at the same time preferring $100 in two years over $50 in a year (Ainslie, 2001).

The conflict theory has also inspired dual-systems models of cognition. System 1 is described as intuitive, associationist, resource-independent, whereas System 2 as deliberative, rule-driven, and resource-dependent. System 1 comprises the features characteristic of irrational (or, at least, non-rational) psychological processes, whereas System 2 has the features characteristic of rational processes. Each system is associated with discrete neural substrates (the limbic system and the lateral prefrontal cortex respectively for System 1 and 2; McClure, Ericson, Laibson, Loewenstein, & Cohen, 2007). Yet, for definitions of rationality, psychological processes are secondary to the beliefs they produce. Beliefs are incoherent if they comprise outright contradictions, such as preference reversals (as in hyperbolic discountin), violations of deductive logic, or impossible probability estimates (Dawes, 1998). The conflict theory attributes the achievement of coherence to System 2 and its breakdown to System 1, but there are exceptions. For example, the reduction of cognitive dissonance has long been regarded as a prototype of irrational reasoning. However, once dissonance is reduced, beliefs are coherent, and the reduction can occur automatically, that is, by System 1 thinking.

The coherence criterion is attractive because of its simplicity and generality. It subsumes other definitional criteria, such as self-interest. Experience and experiments suggest that humans can be made to do what they would rather not do. Compliance, as obtained for example by the "foot-in-the-door" technique, is defined as a behavior that is elicited against the person's resistance and better judgment. Although the person may not be experiencing a psychological conflict, the behavior is irrational in the sense that it violates self-interest, or more broadly, in that the behavior does not cohere with the person's preferences.

Although incoherence is a necessary condition of irrationality, it is over-inclusive. Consider framing effects. When faced with a choice between certain outcome and a risky outcome with the same expected value, they tend to be risk-averse only when the certain outcome is a gain, not when it is a loss (Tversky & Kahneman, 1984). The frame does not affect the expected values of the prospects, and should be ignored. Although incoherent as a pattern, the individual decisions can be justified. Framing effects are akin to the perception of multi-stable figures. When faced with the Necker cube, the perceptual system fails the coherence test because its task is underspecified. Providing alternate interpretations of an ambiguous display is an intelligent solution. Likewise, if two options have the same expected value, there is insufficient reason to consistently choose one over the other. Further problems arise when more than one coherence criterion may apply. Certain varieties of the sunk-cost effect make it possible for a person to be coherent and incoherent at the same time. Someone who continues to invest in a venture that is known to fail incoherently chooses behavior that contradicts monetary self-interest. Yet, this person can point to the behavior's consistency over time.

In some contexts, the rationality of belief is not aligned with the rationality of behavior. Many gamblers believe that an acknowledged random process will correct itself after a run. When a fair die yields for odds numbers in a row, they believe an even number is due. This belief is irrational because it is false and the gambler should know better. Yet a change in the betting behavior is not irrational because the probability of winning is still the same. Various superstitions and magical beliefs are similar in that their epistemology is irrational. The idea that guardian angels protected the accident survivor but forsook the dead victim is untestable, *post hoc*, and unparsimonious (it adds nothing). A common defense of such beliefs is that they please the believer and thus serve self-interest. Like the gambler, however, the esoteric believer also expects tangible gains. The question is how much the believer will pay for the maintenance of these ideas (suppose the guardian angels—or their clever earthly agents—charge for each prayer). Paying money for nothing is irrational.

Theories of rationality have played a unique role in cognitive psychology because they provide a context for the evaluation of human performance. According to one school of thought, these theories are useful to the extent that people's actual decisions and choices depart from theoretical norms. These departures are thought to reveal the mental processes underlying human judgment and decision-making. According to a different school, the same departures reveal the inadequacies of the norms, and the goal of research is to uncover the conditions under which human performance—given the constraints of the underlying mental architecture—is optimal. The debate over the criteria of rationality will continue, and so will the debate over whether humans are *basically* rational or not. Pragmatically, the greatest value of the concept of rationality is that it is a powerful engine driving research into the processes underlying judgment and decision-making.

Word count: 971

References

Ainslie, G. (2001). *Breakdown of will*. Cambridge, MA: Cambridge University Press.

Dawes, R. M. (1998). Behavioral decision making and judgment. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vol. 1, pp 497-548). New York, NY: McGraw-Hill.

McClure, S. M., Ericson, K. M., Laibson, D. I., Loewenstein, G., & Cohen, J. D. (2007). Time discounting for primary rewards. *The Journal of Neuroscience, 23*, 5796-5804.

Tversky, A., & Kahneman, D. (1984). Choices, values, and frames. *American Psychologist, 39*, 341-350.

Suggested Readings

Gintis, H. (2007). A framework for the unification of the behavioral sciences. *Behavioral and Brain Sciences, 30*, 1-61.

Krueger, J. I., & Funder, D. C. (2004). Towards a balanced social psychology: Causes, consequences and cures for the problem-seeking approach to social behavior and cognition. *Behavioral and Brain Sciences, 27*, 313-376.

Keywords

coherence, logic, superstition, dual-systems theories