MIT Press • Open Encyclopedia of Cognitive Science

Human Rationality in Research

Ulrike Hahn[1]

[1]Birkbeck College, University of London

MIT Press

Published on: Jun 17, 2026

DOI: https://doi.org/10.21428/e2759450.84dbfd75

License: Creative Commons Attribution-NonCommercial 4.0 International License (CC-BY-NC 4.0)

Rationality—that is, reasonable judgment, evaluation of evidence, and decision-making—is a topic that runs through cognitive science both on a methodological level and as a topic of empirical inquiry. Over millennia, research has developed normative standards against which the rationality of human and nonhuman agents is evaluated. Different applications of those normative standards within cognitive science include the method of rational analysis and the use of decision theory as a conceptual framework for cognitive neuroscience that goes beyond mere stimulus–response relationships. At the same time, a wealth of empirical research has sought to determine the extent to which human and nonhuman agents (descriptively) live up to these rational norms.

History

The core standards for evaluating human rationality (the "yardsticks" against which behavior is measured) vary substantially in age; logic has its roots in antiquity, probability theory and decision theory reach back to the 17th century, game theory originates in the early 20th century, and attempts to understand rationality more loosely as the adaptive behavior of resource-bounded agents (Simon, 1957) are even more recent. In other words, the frameworks that collectively seek to define rationality have been developed bit by bit. These developments increasingly cover more and more aspects of behavior. However, there are also inclusion relationships between these frameworks; errors in classical logic are also violations of probability theory, probability is a component of decision theory, and game theory typically assumes that agents are individually rational in the sense that decision theory prescribes [see Foundations of Rationality for an introduction to specific normative standards; see Normativity for background on the notion of normativity itself].

In the history of the cognitive sciences, the diversity of these standards is matched by the fact that each of these normative standards ("yardsticks") has given rise to waves of empirical research that have waxed and waned, varying also with respect to the fields and disciplines that have been the primary drivers of that research. There have also been repeated shifts in the extent to which empirical research across these different areas has been interconnected or has proceeded in comparative isolation. There is consequently not a single research area, or homogenous field, empirically examining rationality (see Hahn, 2026 for a more in-depth discussion).

One way to structure the resultant diversity of empirical findings would be by research questions. The different core standards are historically associated with different tasks such as drawing inferences ("There's smoke in the air, there must be a fire") as opposed to making decisions ("To douse the fire, do I choose the water bucket or the blanket?"). However, as formal frameworks, these normative standards are incredibly general. As a consequence, many "tasks" that are formally the same from the perspective of these frameworks are so different in other respects that empirical researchers would typically not consider them to be the same, nor necessarily even related, in terms of underlying cognitive process. For example, we would not expect much process overlap between putatively Bayesian inference automatically conducted by the perceptual system over the course of seconds and Bayesian inference conducted by participants in a cognitive, scenario-based task. It is consequently easier to think about the breadth of rationality-focused research from the perspective of the different frameworks themselves [see Bayesian Models of Cognition].

Core concepts

Logic

Classical logic concerns itself with inferences that follow by necessity (such that if the premises are true, the conclusions are necessarily true as well). Experimental studies examining the extent to which human reasoning lives up to the standards of logic have typically focused on *syllogisms* (argument schemes in which a conclusion is drawn from two or more premises; see, e.g., Johnson-Laird & Bara, 1984) and on reasoning with conditionals (propositions that can be cast as "if-then" statements; Wason, 1968). The bulk of this work has taken place in the psychology of reasoning. However, that area of research has been reshaped by the evidence that performance on classical reasoning tasks might be better understood as participants engaging in probabilistic reasoning, which has given rise to a *probabilistic turn* in that field (Oaksford & Chater, 2020).

Probability

Probability theory is a formal framework for capturing uncertainty and reasoning about things under conditions of uncertainty. Judgment and decision-making research in psychology has accumulated a wealth of empirical research on ways in which everyday estimates of probability might go wrong, including ways in which these might systematically be biased. This is supplemented by bias-focused research in other areas such as social psychology (Krueger & Funder, 2004) and behavioral economics. With Bayes' rule, probability theory also includes normative standards for belief revision in light of new information [see Bayesianism]. This has prompted research into evidential reasoning (Phillips & Edwards, 1966), including in specialist domains such as law (Fenton et al., 2013), and ways in which such reasoning might be improved (Gigerenzer & Hoffrage, 1995). The generality and flexibility of probability theory as a framework has also provided the basis for rationality-themed research in causal reasoning, argumentation, and critical thinking (Hahn, 2020) [see Causal Reasoning].

Decision theory and game theory

Decision theory concerns itself with choosing between options when these involve uncertainty. Making such decisions in line with decision theory requires attending (only) to the probabilities associated with different possible outcomes and to the (subjective) value each of these outcomes holds. This generality means that empirical work is even more diverse than it is in the case of probability. One can study the rationality of decisions as varied as buying insurance, deciding whether to conduct a medical test, or eye movements in scanning the visual environment. Because decision theory defines the rational actor model that is a core building block of much economic theory, empirical research on its descriptive accuracy has been a core topic not just for psychology but also for economics. At the same time, the use of decision theory as an analytical framework in cognitive neuroscience (Glimcher, 2004) has generated research on the optimality of many different perceptual (Navalpakkam et al., 2010) and perceptuomotor decisions (e.g., Trommershäuser et al., 2008) [see Neuroeconomics]. Game theory finally concerns itself with strategic decisions in which an agent's best course of action depends also on what other agents choose (Davis, 1997). A core topic that has attracted a wealth of empirical attention is the fact that people tend to behave more altruistically and cooperatively than the prescriptions of game theory suggest (Kollock, 1998) [see Economic Games].

Adaptive rationality

Cutting across these different areas is the perspective that these core norms of rationality may be overly demanding and that seeming failures might reflect the use of simple but adaptive heuristics (Simon, 1957). Some of the most extensively studied violations of rational norms such as the *conjunction fallacy* (in which the conjunction of two statements, A and B, is rated as more probable than its respective components; see Tversky & Kahneman, 1983) or *base rate neglect* (in which the base rate is given insufficient weight in determining the impact of evidence; see Kahneman & Tversky, 1973) were originally linked to the use of heuristics. Those heuristics (such as *representativeness* or *availability*) themselves came under intense scrutiny (both empirically and conceptually), and subsequent empirical programs aimed at identifying useful heuristics sought to place greater emphasis on the fact that adaptation means adaptation to an environment. Consequently, the success or failure of a given heuristic rests on the specific environment, and the characterization of the relevant environment therefore needs to be a core part of any analysis in terms of adaptive rationality (Gigerenzer et al., 1999).

Questions, controversies, and new developments

Two strands should be highlighted with respect to open questions and further developments. First, current norms of rationality still seem incomplete in that there are many additional questions that feel like they should be subject to rational constraints but are not covered (or covered adequately) by extant normative frameworks. So, frameworks of rationality are still very much under development and this may mean further changes to the range of behaviors under empirical scrutiny.

Second, important questions remain concerning the applicability of these frameworks beyond individual human agents. There are empirical studies of rational decision-making in nonhuman species, nonhuman primates in particular (Hurley & Nudds, 2006), but does it equally make sense to apply decision theory to an organism like slime mold (e.g., Latty & Beekman, 2011)? Not only is slime mold a brainless organism far below the cognitive capacity we associate with reasoning or decision-making, it is the collective that, if anything, is making decisions. The collective dimension is already in play in thinking about the rationality of human groups and whether rational collectives may arise from irrational individuals and vice versa (Mayo-Wilson et al., 2011) [see Social Epistemology].

The question of requisite cognitive capacity is also in play with respect to computational agents such as generative artificial intelligence (AI) systems (Macmillan-Scott & Musolesi, 2024) [see Large Language Models]. There is an intuition that associates rationality with emotionless, cold calculation that might make machines seem rational by definition. However, the experimental literature on rationality contains a wealth of research on the relationship between rationality and emotion (for an overview, see Pham, 2007) that does not impute such a rigid divide [see Foundations of Emotion]. Furthermore, although it is a largely conceptual issue whether AI systems should be credited with reasoning or with decision-making, it is a wholly empirical question whether their behavior is in line with the constraints imposed by the normative frameworks of rationality. Research must empirically assess whether AI systems' outputs are contradictory (violating the constraints of logic), whether the options they propose maximize expected utility in the way decision theory demands (Binz & Schulz, 2023), and how they fare in game theoretic contexts (Akata et al., 2025).

Broader connections

The very concept of rationality has always been tied up with the question of what it means to be human. In that sense, it is no surprise that not just experimental data, but also the theoretical conception of rationality itself, have been challenged with respect to the extent to which they genuinely reflect human experience in general. Not only are participant populations overwhelmingly WEIRD (Western, Educated, Industrialized, Rich, and Democratic), there is evidence to suggest that this may affect characteristics of particular relevance to empirical studies of rationality (such as risk attitudes or altruism) [see WEIRD]. On a conceptual level, feminist philosophy, in particular, has questioned the universality of notions of reasonableness captured by norms of rationality (see e.g., Lloyd, 2017). At the same time, the standard normative frameworks play a central role in economics, epistemology, and the philosophy of science. Empirical research on rationality thus sits at the intersection of many debates that reach beyond the attempt to characterize patterns of human behavior.

Further reading

- Gigerenzer, G. (2025). The rationality wars: A personal reflection. *Behavioural Public Policy*, *9*(3), 495-515. https://doi.org/10.1017/bpp.2024.51
- Hahn, U. (2026). Human rationality. *Annual Review of Psychology*, *77*, 193-221. https://doi.org/10.1146/annurev-psych-020425-020958

References

- Akata, E., Schulz, L., Coda-Forno, J., Oh, S. J., Bethge, M., & Schulz, E. (2025). Playing repeated games with large language models. *Nature Human Behaviour*, *9*(7), 1380-1390. https://doi.org/10.1038/s41562-025-02172-y

↩

- Binz, M., & Schulz, E. (2023). Using cognitive psychology to understand GPT-3. *Proceedings of the National Academy of Sciences*, *120*(6), e2218523120. https://doi.org/10.1073/pnas.2218523120

  ↩
- Davis, M. D. (1997). *Game theory: A nontechnical introduction*. Dover Publications.

  ↩
- Fenton, N., Neil, M., & Lagnado, D. A. (2013). A general structure for legal arguments about evidence using Bayesian networks. *Cognitive Science*, *37*(1), 61-102. https://doi.org/10.1111/cogs.12004

  ↩
- Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, *102*(4), 684-704. https://doi.org/10.1037/0033-295X.102.4.684

  ↩
- Gigerenzer, G., Todd, P. M., & The ABC Research Group. (1999). *Simple heuristics that make us smart*. Oxford University Press.

  ↩
- Glimcher, P. W. (2004). *Decisions, uncertainty, and the brain: The science of neuroeconomics*. MIT Press.

  ↩
- Hahn, U. (2020). Argument quality in real world argumentation. *Trends in Cognitive Sciences*, *24*(5), 363-374. https://doi.org/10.1016/j.tics.2020.01.004

  ↩
- Hahn, U. (2026). Human rationality. *Annual Review of Psychology*, *77*, 193-221. https://doi.org/10.1146/annurev-psych-020425-020958

  ↩
- Hurley, S., & Nudds, M. (Eds.). (2006). *Rational animals?* Oxford University Press.

  ↩
- Johnson-Laird, P. N., & Bara, B. G. (1984). Syllogistic inference. *Cognition*, *16*(1), 1-61. https://doi.org/10.1016/0010-0277(84)90035-0

  ↩
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*(4), 237-251. https://doi.org/10.1037/h0034747

  ↩

- Kollock, P. (1998). Social dilemmas: The anatomy of cooperation. *Annual Review of Sociology*, *24*(1), 183-214. https://doi.org/10.1146/annurev.soc.24.1.183

↩

- 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*(3), 313–327. https://doi.org/10.1017/s0140525x04000081

↩

- Latty, T., & Beekman, M. (2011). Speed–accuracy trade-offs during foraging decisions in the acellular slime mould *Physarum polycephalum*. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1705), 539-545. https://doi.org/10.1098/rspb.2010.1624

↩

- Lloyd, G. (2017). Rationality. In A. M. Jaggar & I. M. Young (Eds.), *A companion to feminist philosophy* (pp. 163-172). Wiley Blackwell.

↩

- Macmillan-Scott, O., & Musolesi, M. (2024). (Ir)rationality and cognitive biases in large language models. *Royal Society Open Science*, *11*(6), 240255. https://doi.org/10.1098/rsos.240255

↩

- Mayo-Wilson, C., Zollman, K. J., & Danks, D. (2011). The independence thesis: When individual and social epistemology diverge. *Philosophy of Science*, *78*(4), 653-677. https://doi.org/10.1086/661777

↩

- Navalpakkam, V., Koch, C., Rangel, A., & Perona, P. (2010). Optimal reward harvesting in complex perceptual environments. *Proceedings of the National Academy of Sciences*, *107*(11), 5232-5237. https://doi.org/10.1073/pnas.0911972107

↩

- Oaksford, M., & Chater, N. (2020). New paradigms in the psychology of reasoning. *Annual Review of Psychology*, *71*(1), 305-330. https://doi.org/10.1146/annurev-psych-010419-051132

↩

- Pham, M. T. (2007). Emotion and rationality: A critical review and interpretation of empirical evidence. *Review of General Psychology*, *11*(2), 155-178. https://doi.org/10.1037/1089-2680.11.2.155

↩

- Phillips, L., & Edwards, W. (1966). Conservatism in a simple probability inference task. *Journal of Experimental Psychology*, *72*(3), 346–354. https://doi.org/10.1037/h0023653

↩

- Simon, H. A. (1957). *Models of man, social and rational: Mathematical essays on rational human behavior in a social setting.* Wiley.

  ↩
- Trommershäuser, J., Maloney, L. T., & Landy, M. S. (2008). Decision making, movement planning and statistical decision theory. *Trends in Cognitive Sciences, 12*(8), 291-297. https://doi.org/10.1016/j.tics.2008.04.010

  ↩
- Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*(4), 293-315. https://doi.org/10.1037/0033-295X.90.4.293

  ↩
- Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, *20*(3), 273-281. https://doi.org/10.1080/14640746808400161

  ↩