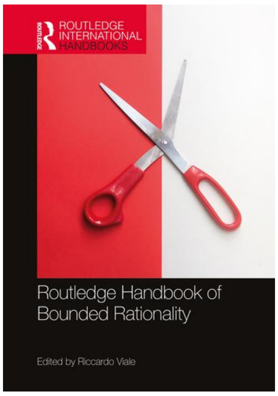

Routledge Handbook of Bounded Rationality

Riccardo Viale

How rationality is bounded by the brain

Publication details
https://test.routledgehandbooks.com/doi/10.4324/9781315658353-29

Paul Thagard

How to cite :- Paul Thagard. 29 Oct 2020, *How rationality is bounded by the brain from:* Routledge Handbook of Bounded Rationality Routledge
Accessed on: 30 Mar 2023
https://test.routledgehandbooks.com/doi/10.4324/9781315658353-29

25

HOW RATIONALITY IS BOUNDED BY THE BRAIN

Paul Thagard

When Richard Thaler was asked how he planned to spend the million dollars from his 2017 Nobel Prize in economics, he replied: “I will try to spend it as irrationally as possible” (Politi, 2017). Thaler’s prize was awarded for decades of research showing that people are not nearly as rational as economists have assumed.

Why are people frequently irrational? There are still many economists and philosophers who believe that people are fundamentally rational, but a large accumulation of evidence from psychology and behavioral economics shows that people often fall short of rational standards. The evidence is descriptive, showing that people make thinking errors in systematic ways. But these descriptions leave open the question of *why* people do not think in ways that support their long-term interests.

Cognitive science usually explains thinking in terms of mental representations and processes, but advances in neuroscience make it increasingly feasible to explain many mental processes as brain mechanisms. So the question becomes: What are the brain mechanisms that lead people to be irrational?

The term “bounded rationality” originated in the 1950s in the writings of Herbert Simon (e.g., Simon, 1972), but appreciation that people are often limited in their rationality goes back to Aristotle and Francis Bacon. The human brain is marvelous in many of its accomplishments, but I will describe its inherent limitations in size, speed, and cognitive-emotional functioning.

Herbert Simon (2000, p. 25) characterized bounded rationality as follows:

> Bounded rationality is simply the idea that the choices people make are determined not only by some consistent overall goal and the properties of the external world, but also by the knowledge that decision makers do and don't have of the world, their ability or inability to evoke that knowledge when it is relevant, to work out the consequences of their actions, to conjure up possible courses of action, to cope with uncertainty (including uncertainty deriving from the possible responses of other actors), and to adjudicate among their many competing wants. Rationality is bounded because these abilities are severely limited.

I show how these mental limitations derive in part from brain limitations.

Table 25.1 3-analysis of *rational*

Exemplars	Logical deduction, probabilistic reasoning and statistical inference, making decisions based on maximizing expected utility.
Typical features	Following rules of reason, careful, conscious.
Explanations	Explains: why people get true beliefs; why people make good decisions. Explained by: ability of humans to follow appropriate rules of reasoning.

Table 25.2 3-analysis of *irrational*

	Fallacious deductive inferences such as affirming the consequent; defective probabilistic inferences such as judging a conjunction to be more probable than either of its conjuncts; making bad decisions such as emphasizing sunk costs rather than future expectations.
Typical features	Violation of normative rules, careless succumbing to fallacies and biases.
Explanations	Explains: why people arrive at dumb beliefs and make bad decisions. Explained by: susceptibility to psychological processes that interfere with the application of good rules.

What is rationality?

The question "what is rationality?" sounds like a request for a definition, but decades of work on the psychology of concepts show that concepts outside of mathematics are rarely susceptible to clear definitions in terms of necessary and sufficient conditions. Rather, concepts are better characterized in terms of the exemplars that provide standard examples of them, typical conditions for the concept, and explanations that the concepts can be used to provide (Murphy, 2002). A new neural theory of concepts shows how to combine all three of these aspects of concepts in a unified neural model (Blouw, Solodkin, Thagard, & Eliasmith, 2016).

Accordingly, we can characterize concepts by identifying these three dimensions of exemplars, typical features, and explanations, a method of conceptual analysis that Thagard (2019a) calls "3-analysis." Table 25.1 provides a 3-analysis of the concept *rational*.

According to currently dominant traditions, standard examples of rationality include using formal logic to deduce truths from truths, using probability theory to perform inductive reasoning, and using maximization of expected utility to decide what to do by combining probabilities and utilities to choose actions. The typical features of being rational include carefully and consciously applying such normative rules. Such applications explain why people sometimes succeed in acquiring true beliefs and making good decisions. That people are rational is explained by their ability to follow rules. This concept of rationality is complemented by the prevalent concept of irrationality, which can also be captured by a 3-analysis, as shown in Table 25.2.

The recognition of irrationality

Formal logic begin with Aristotle's doctrine of the syllogism, but Aristotle also recognized that people often fall short of good syllogistic reasoning. His *Sophistical Refutations* describes numerous fallacies that people are prone to commit, such as equivocating on the meanings of ambiguous words (Aristotle, 1984). The study of fallacious reasoning has a long history in philosophy and survives today in the discipline of informal logic (Hanson, 2015).

In the seventeenth century, Francis Bacon (1960) provided a sophisticated discussion of inductive reasoning in *Novum Organon*. In addition to giving good advice about how to go from observations to generalizations, Bacon generated a list of mistakes that people often make in inductive reasoning. He called them "idols," which he colorfully described as idols of the tribe (due to human nature), idols of the cave (due to what a particular human cares about), idols of the marketplace (due to communication), and idols of the theater (due to philosophical prejudices). Bacon's idols capture some of the errors recognized by twentieth-century psychologists, such as the availability heuristic and motivated inference.

In the 1950s, Herbert Simon recognized the limited extent of human rationality, but he did not systematically investigate the ways in which people fall short of good decision making and inductive inference. Beginning in the 1970s, psychologists such as Daniel Kahneman and Amos Tversky and economists, such as Richard Thaler, conducted experiments that identify many ways in which human thinking falls short of good reasoning (Kahneman & Tversky, 2000; Thaler, 2015).

I use the term "error tendencies" to cover all the fallacies, idols, heuristics, biases, and typical mistakes that have been identified by philosophers, psychologists, and economists, and have compiled more than 50 of them (Thagard, 2011). Why are there so many? Why did evolution by natural selection fail to optimize people's ability to reason well about what to believe and what to do?

The biological answer is that brains have evolved to be only somewhat effective at performing deductive, inductive, and practical inferences. Optimization is constrained by biological mechanisms and the difficulties of survival and reproduction in changing environments. I will describe how these constraints have produced brain mechanisms that often work well but are limited by size, speed, cognitive-emotional functions, and limitations of attention and consciousness.

Brain size and speed

The argument that people must be rational because of optimization through natural selection has two flaws. First, it gets evolution wrong, because natural selection does not optimize (Gould & Lewontin, 1979). Rather, nature selects for organisms that are somewhat better at surviving and reproducing than organisms with different genes.

Second, the optimization argument forgets that the current standards of rationality are relatively recent cultural innovations. Humans have been around for at least 100,000 years, but formal logic only began with Aristotle around 2,500 years ago, and sophisticated understanding of deduction only began with the work of Gottlob Frege and Charles Peirce in the late nineteenth century. Probability and utility theories are products of mathematical thinking that began in the seventeenth and eighteenth centuries, and the elegant version now used by economists was only developed in the 1940s. For some purposes, these tools are useful, but there is no reason to suppose that they are built into the human brain by evolution.

Human brains have numerous strengths that have enabled people to spread all over the planet and increase in population to more than 7 billion. The most impressive features of human brains are not the special-purpose adaptations touted by evolutionary psychologists, but rather the general adaptability furnished by the flexible ways in which humans can learn from experience (Quartz & Sejnowski, 1997).

Nevertheless, the brain has numerous limitations that forestall optimal rationality. Our brain's assemblage of 86 billion neurons provides a lot of computing power, but elephants have three times as many. In order to have more neurons, people would need to have bigger brains that require bigger heads, but childbirth is already often a difficult procedure. Human brain size reflects a trade-off between the benefits of more processing power and ease of delivery through a pelvis that also must function for bipedal locomotion.

Another constraint on human brain size concerns energy. Even though the roughly 1.4 kg of the human brain take up less than 3 percent of the average human weight, the brain uses up to 20 percent of the energy available to the body. Larger brains would require more energy, which either requires less energy available for other functions such as metabolism and reproduction, or greater sources of food. The evolution of human brains requires a trade-off between size and energy efficiency, as well as between size and birth delivery.

These limitations on the size of human brains place important constraints on rationality because there is a limited amount of information that people can store. Some philosophers have maintained that it is rational to believe all the logical consequences of one's beliefs. An infinite set cannot be stored in any human brain or even in all the computers run by Amazon and Google.

Moreover, even for a finite number of beliefs, the size limitation of the brain places sharp constraints on the combinations of beliefs that can be considered. Educated people have a vocabulary of around 30,000 words, so with 10 beliefs for each word they would have 300,000 beliefs. Understanding is growing of how such beliefs can be stored in the human brain through distributed representations (Eliasmith, 2013). But brains cannot accommodate belief revisions that require consideration of combinations of $2^{300,000}$ subsets of these beliefs. Such subsets are required for considering whether human belief systems are consistent. So it is unreasonable to expect that people should be consistent in their beliefs.

Human brains also come with limitations in speed of processing. Billions of neurons allow for massively parallel operation, but the neurons themselves are slow. A typical neuron fires up to 200 times per second, whereas current computers have operations at the rate of trillions of times per second.

Why are neurons so slow? Most neural connections are chemical, requiring the movement of neurotransmitters such as glutamate and gamma aminobutyric acid (GABA) from one neuron to another. This chemical transmission is slower than purely electrical signaling, which occurs rarely in brains, but it allows for flexibility in timing and the development of different kinds of pathways.

In principle, the brain could operate with only two neurotransmitters, one for enabling one neuron to excite another (increasing its rate of firing), and another for enabling one neuron to inhibit another (decreasing its rate of firing). Glutamate and GABA play these excitatory and inhibitory roles, respectively. But there also dozens of neurotransmitters that operate in human brains, carrying out diverse functions at different time scales. For example, circuitry involving the neurotransmitter dopamine is important for motor control and learning about rewards, while circuitry involving serotonin influences perception and emotion.

If brains were faster, they still would not be able to do an infinite amount of processing, but they would be able to better approximate some of the standards required for the rational norms of deductive logic and probability and utility theory. The brain lacks the speed to be able to do all of the calculations that are required for absolute standards of rationality. I estimate that speed and size limitations contribute to more than 30 of the 53 error tendencies listed in Thagard (2011). For example, people would be less prone to representativeness (the tendency to use assessments of similarity in causal reasoning) if they had the cognitive capacity to do fuller statistical calculations.

On the other hand, brains are appropriately efficient at carrying out computations that are important for the survival and reproduction of organisms. Perception, inference, and decision-making can all be modeled as processes of parallel constraint satisfaction, in which a brain or computer considers a range of possible interpretations of a conflict situation and comes up with a good but not necessarily optimal solution (Rumelhart & McClelland, 1986; Thagard, 2019a). For example, recognizing a moving object as an instance of prey or predator should consider alternative hypotheses about the animal, constrained by perceptual and environmental information. Parallel constraint satisfaction is efficiently computed by neural networks that implement constraints by excitatory and inhibitory links (Thagard, 2000).

Therefore, the brain can be understood as an engine of coherence rather than deduction or calculation. Coherence requires satisfying multiple constraints in parallel, not making zillions of calculations. In general, coherence is computationally intractable (Thagard and Verbeurgt, 1998), but in practice is efficient as long as the number of inhibitory connections is small compared to the number of excitatory ones, as is true of the brain (van Rooij, Blokpoel, Kwishout, & Wareham, 2018).

In sum, human rationality is bounded by the limited size and speed of the brain, but it can often function well by making coherence judgments rather than overextending itself by making too many deductions and calculations. In exciting work in progress, Dan Simon and Stephen Read show that a large number of cognitive biases can be explained by thinking of the mind as a coherence process. My aim is less specific, to show why the alternative methods of rational calculation are at odds with the size and speed of the brain.

Brain integration of cognition and emotion

Rationality is also bounded by more specific aspects of how brains function, including the integration of cognition and emotion and the limited role of consciousness. An obsolete view of the brain takes it as combining a recent cognitive system consisting of areas such as the prefrontal cortex built on top of an ancient limbic system consisting of primitive areas such as the amygdala. This view was exploded by findings that the most high-level areas are intensely interconnected with the emotional systems (Damasio, 1994; Pessoa, 2013).

The integration of cognition and emotion in the brain is generally a feature rather than a bug. Accounts of deduction and probability assume that the brain is largely a syntactic engine, with semantics (meaning) and pragmatics (context and purpose) only playing peripheral roles. But syntax alone cannot explain numerous aspects of human thinking, such as performance in the selection task of Wason (1966) and the complexity of analogical inference (Holyoak & Thagard, 1995). Standard computer programs are fabulous at rapid syntactic manipulations, but are much less successful in the semantics of connections to the world and the pragmatics of accomplishing important goals in particular contexts.

In contrast, neural representations mingle syntax and semantics by means of representations that can simultaneously handle relational structure and connections to the world. The best current account of this mingling is Chris Eliasmith's (2013) Semantic Pointer Architecture, which shows how populations of neurons can retain information gained from the world by sensory and motor operations, but also build up syntactically-rich structures. It thereby provides a biologically plausible synthesis of syntax and semantics.

This architecture extends to explain emotions as brain representations that integrate information drawn from (1) physiological changes in the body; (2) cognitive appraisals in the brain; and (3) contextual information including the use of language in humans (Thagard & Schröder, 2014; Thagard 2019a, 2019b; Kajić, Schröder, Stewart, & Thagard, 2019). Emotions that pervade cognition show how goals and purposes can fundamentally influence the operation of the neural system, making it a pragmatic as well as a syntactic and semantic addition. Hence adding emotions to the Semantic Pointer Architecture shows how brains accomplish a synthesis of syntax, semantics, and pragmatics. This integration helps to ensure that human thinking is not just idle deduction or calculation, but operates effectively in the world and accomplishes human goals and purposes.

Nevertheless, the evolutionary feature of integration of cognition and emotion comes with bugs. In the brain, cognitions and emotions are not independent of each other, unlike the calculations in expected utility theory where probabilities and utilities are distinct before being mathematically combined. Under different circumstances, the brain might have evolved with separate modules for probability calculations, utility calculations, and their integration in calculations of expected utility. But these are cultural developments that came late in the history of our species.

Some of the error tendencies (bugs) that arise from illicit mingling of cognitions and emotions are shown in Table 25.3. Normatively, calculations of probabilities and utilities should be independent of each other, but independence fails in the brain. Motivated inference is the well-known tendency of people to base their beliefs not just on relevant evidence, but on what they want to believe (Kunda, 1990). This thinking is more complicated than wishful thinking, because it requires interactions of goals, memory, and inference making. The brain has no firewall between cognition and emotion, so it is not surprising that people often adopt beliefs that they find emotionally appealing, in domains that range from politics to relationships. For example, people who like particular politicians find it hard to believe that they have misbehaved.

Surprisingly, however, people do not always believe what makes them happy, but instead believe things because they scare them, which is fear-driven inference. The classic example is Shakespeare's Othello, who has only scanty evidence that his wife is unfaithful, but cannot block the inference of infidelity because fear keeps him thinking about it. The emotional significance of the hypothesis of unfaithfulness is so strong that it hijacks attention while Othello ignores contrary evidence and alternative hypotheses. High disutility prompts an estimation of high probability.

Another illicit interaction of probability and utility is sour grapes, where the fox concludes that the grapes are sour because he cannot reach them. Logically, low utility should be

Table 25.3 Error tendencies resulting from confusions of probability and utility caused by cognition-emotion interactions

Error tendency	*Example*	*Illicit interaction*
Motivated inference	I really want that job, so I'm sure to get it.	high utility → high probability
Fear-driven inference	I'm terrified that this mole is cancerous, so it must be.	high disutility → high probability
Sour grapes	I can't afford a BMW, but they're too unreliable anyway.	low probability → low utility
Rationalization	I have to go to Moose Jaw, so it should be a fun city.	high probability → high utility

Source: Based on Thagard (2019c).

independent of low probability, but the brain again mixes cognition and emotion to yield the dubious conclusion.

Another mixture that people succumb to is rationalization, where the high probability of an occurrence makes us think that it is at least okay, with more utility than we would otherwise judge it to have. After Donald Trump was elected president, many pundits erroneously argued that it could not really bad as initially feared because of political constraints on the presidency. For emotional rather than cognitive reasons, people have a tendency to accept their current situations.

Other familiar error tendencies in decision making result from inappropriate interconnections of cognition and emotion. People are prone to the fallacy of sunk costs, making their decisions based on past results rather than on future expectations. For example, people sometimes stay in careers and romantic relationships too long, because of past investments rather than considerations of future prospects. Instead of calculating the expected utility of continuing the career or the relationship, people are driven to avoid the emotion of regret that they would feel if they bailed out, along with the possibility of other negative social emotions such as embarrassment, guilt, and shame. Such emotions get in the way of rational calculations about what to do in the future rather than focusing on the unchangeable past.

Similarly, emotional effects on cognition explain the well-known tendency of people to be unduly influenced by immediate context rather than by long-term effects. People have a tendency to go for short-term small gains in neglect of long-term large gains that they prefer on reflection. For example, people tend to buy things immediately rather than save for retirement. A famous example of this time discounting is the psychological study in which most children choose to eat a marshmallow immediately rather than wait a short time to get two marshmallows.

The neural explanation of time discounting is that different brain areas are involved (McClure et al., 2007). Decisions about what to do immediately engage emotion-related areas such as the ventral striatum and orbitofrontal cortex, whereas decisions about what to do in the long run are not so emotionally engaging and therefore can be done by calculations in prefrontal and parietal areas. Faced with immediate rewards, people do not perform calculations of long-term expected utilities, with emotion overwhelming cognition.

Other well-known incursions of emotion into decision making include risk aversion and the tendency to frame losses as more salient than gains. Neurobiological investigations find that loss aversion correlates with activity in the amygdala, suggesting that dealing with losses has greater emotional effect than dealing with gains (De Martino, Kumaran, Seymour, & Dolan, 2006). I estimate that more than half of the 53 error tendencies listed in Thagard (2011) have a substantial emotional component. Hence the integration of cognition and emotion in the brain contributes as much as size and speed limitations to the boundedness of rationality.

Brain limitations on attention and consciousness

In his best-selling book, *Thinking, Fast and Slow*, Daniel Kahneman (2011) uses dual processes theories of mind to explain why people are so prone to the many thinking biases that he and Amos Tversky identified. Like many other psychologists, he distinguishes between System 1, which is fast, automatic, involuntary, and unconscious, and System 2 which is slow, deliberate, voluntary, and conscious. The distinction explains numerous thinking biases as resulting from the unreflective operations of System 1, whereas System 2 allows people to apply appropriate formal rules such as theories of probability and utility.

The problem with this explanation is that there is no specification of how the two systems actually operate. Suggestions have been made about how the two systems might map onto different areas of the brain, but these proposals have not stood up to empirical scrutiny (Spunt, 2015). Dual process theory is a useful, descriptive, way of classifying different kinds of thinking, but provides no explanation of them, because it does not specify the mechanisms that underlie either of the processes.

Thagard (2019a, Chapter 8) uses the Semantic Pointer Architecture to describe mechanisms that produce the differences between thinking fast and slow. In the slow mode, inferences and actions take place because of interactions among brain areas that interpret sensory inputs, evaluate sensory inputs, and generate new neural representations. The slow mode is much rarer and occurs when competition among brain representations generates conscious awareness via a small subset of them.

Since George Miller's (1956) landmark paper on the magical number seven, psychologists have been aware that conscious, working memory is severely limited. The reasons for this limitation are not clear: it may be an adaptive feature designed to focus mental resources on potential actions, which have to be serial rather than parallel; or it may just be a side effect of the large amount of neural resources required to produce bindings of representations accessible to consciousness.

Either way, consciousness is limited in ways that block people's awareness of their failures to follow normative principles rather than committing the errors so far. For example, when people persist in relationships or businesses because of the sunk cost error tendency, they may not be aware that they are failing to do a good calculation of the expected consequences of their actions because of emotions such as regret and embarrassment. On the other hand, the emotional importance of personal relationships and careers should encourage people to consider what past events reveal about the satisfaction of their basic goals, and thereby discourage them from tossing away what worked in the past based on a superficial calculation of future gains. For example, if you have spent years in a romantic relationship, it is worthwhile reflecting on why there were some good times before making inferences about future consequences.

Most inferences in decision-making occur unconsciously, without people being aware of what they were doing. The limits of conscious attention exacerbate the limits of size and speed discussed earlier. Even if the brain has the resources to carry out complex inferences, it often does not have the capacity to become aware of how it is performing. Hence, people cannot consciously check whether they are following appropriating normative rules or just sliding into error tendencies.

Without completely reengineering the brain in a way that is not evolutionarily available, there is no way to enable people to make more of their thinking consciously evaluable by normative principles. People can use external memory such as paper and spreadsheets to write down relevant considerations, overcoming the limitations of working memory and attention.

Conclusion: helping brains to be more rational

Faced with the perplexing plethora of error tendencies, we can ask how people can be helped to think more rationally. The standard pedagogical method is to teach critical thinking courses in which students are made aware of fallacies, biases, and other error tendencies, in the hope that this awareness will reduce mistakes. That is the practice that I followed when I taught critical thinking, and students said they thought that the class helped. But careful studies of the effectiveness of critical thinking instruction are rare (Cotter & Tally, 2009).

A deeper strategy would be to look critically at what counts as rationality. My 3-analysis of rationality in Table 25.1 assumes that people should aim to meet the standards of deductive logic, probability theory, and utility theory, but these have psychological and philosophical limitations. Outside mathematics, logical deduction is rare, so it does not provide much of a standard for rational inference. Far more common is abductive inference, where hypotheses are accepted because they provide the best explanation of the available evidence. Such inference is better understood as based on coherence rather than on formal principles (Thagard, 1989, 2000).

Probability theory is immensely useful when frequencies are known, for example, in games of chance and statistically rich sciences. But the applicability of probabilities understood as subjective degrees of belief is much more contentious, because their potential objectivity and psychological reality are suspect (Thagard, 2019c). When we know statistical probabilities, we should use them, but wildly guessing about them does not further the aims of making good inferences.

Similarly, the theory of expected utility theory assumed by economists rests on psychologically dubious foundations. Whereas nineteenth-century theorists such as Bentham viewed utility as a psychological quantity, twentieth-century economists tried to reconstruct utility from preferences. The result was mathematically elegant, but got the causal explanation backward. People have the preferences that they do because of estimations of value, not utilities because of preferences. Hence it is not clear that the theory of expected utility provides the desired normative standard for decision-making. An alternative is to view decisions as inferences to the best plan, where assessment of actions is based on coherence with emotion-laden goals (Thagard, 2006, 2019c).

Accordingly, we should consider replacing the exemplars of rationality in Table 25.1 with some of the following: coherence-based abductive inference, probabilistic inference in appropriate domains, and decisions based on inference to the best plan (Thagard, 2000). The other dimensions of the 3-analysis survive, but take on a less formal and more realistic aim to help people improve their acquisition of beliefs and their choices of actions. People still qualify as frequently irrational, for example, in motivated inference, but at least we have a better sense of why they succumb because of close ties between normative practices and irrational deviations.

It is an open question whether understanding the roots of bounded rationality in the brain can help people to avoid error tendencies. There need to be controlled experiments that evaluate the effectiveness in improving inferences about beliefs and actions under these conditions:

1. Make people aware of error tendencies with vivid examples, as is currently done in critical thinking courses.
2. Make people aware of error tendencies along with psychological explanations of why people are prone to them, such as dual process theories of cognition and social influences (Nisbett, 2015).
3. Make people aware of error tendencies along with neural explanations of why people are prone to them, including the brain limitations with respect to size, speed, cognition-emotion interactions, and limitations of consciousness.

My conjecture is that understanding why people are so prone to irrational inferences might help students of critical thinking to be more rational.

Regardless of pedagogic effectiveness, understanding how rationality is bounded by the brain should allow psychologists, economists, and philosophers to have a better theoretical understanding of why people are so frequently irrational.

Acknowledgments

The author thanks Riccardo Viale for helpful comments.

References

Aristotle. (1984). *The complete works of Aristotle*. Princeton,, NJ: Princeton University Press.

Bacon, F. (1960). *The New Organon and related writings*. Indianapolis: Bobbs-Merrill.

Blouw, P., Solodkin, E., Thagard, P., & Eliasmith, C. (2016). Concepts as semantic pointers: A framework and computational model. *Cognitive Science*, 40, 1128–1162.

Cotter, E. M., & Tally, C. S. (2009). Do critical thinking exercises improve critical thinking skills? *Educational Research Quarterly*, 33(2), 3.

Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York: G. P. Putnam's Sons.

De Martino, B., Kumaran, D., Seymour, B., & Dolan, R. J. (2006). Frames, biases, and rational decision-making in the human brain. *Science*, 313(5787), 684–687.

Eliasmith, C. (2013). *How to build a brain: A neural architecture for biological cognition*. Oxford: Oxford University Press.

Gould, S. J., & Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Society of London B*, 205(1161), 581–598.

Hanson, H. (2015). Fallacies. *Stanford encyclopedia of philosophy*. Available at: https://plato.stanford.edu/entries/fallacies/

Holyoak, K. J., & Thagard, P. (1995). *Mental leaps: Analogy in creative thought*. Cambridge, MA: MIT Press/Bradford Books.

Kahnemann, D. (2011). *Thinking, fast and slow*. Toronto: Doubleday.

Kahneman, D., & Tversky, A. (Eds.) (2000). *Choices, values, and frames*. Cambridge: Cambridge University Press.

Kajić, I., Schröder, T., Stewart, T. C., & Thagard, P. (2019). The semantic pointer theory of emotions. Cognitive Systems Research, 58: 35–53.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108, 480–498.

McClure, S. M., Ericson, K. M., Laibson, D. I., Loewenstein, G., & Cohen, J. D. (2007). Time discounting for primary rewards. *Journal of Neuroscience*, 27(21), 5796–5804.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81–97.

Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.

Nisbett, R. E. (2015). *Mindware: Tools for smart thinking*. New York: Farrar, Straus, and Giroux.

Pessoa, L. (2013). *The cognitive-emotional brain: From interactions to integration*. Cambridge, MA: MIT Press.

Politi, D. (2017). Richard Thaler wins economics Nobel for recognizing people are irrational. *Slate*. Available at: www.slate.com/blogs/the_slatest/2017/10/09/richard_thaler_wins_economics_nobel_for_recognizing_that_people_are_irrational.html.

Quartz, S. R., & Sejnowski, T. J. (1997). The neural basis of cognitive development: A constructivist manifesto. *Behavioral and Brain Sciences*, 20, 537–556.

Rumelhart, D. E., & McClelland, J. L. (Eds.) (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*. Cambridge, MA: MIT Press/Bradford Books.

Simon, H. A. (1972). Theories of bounded rationality. *Decision and Organization*, 1(1), 161–176.

Simon, H. A. (2000). Bounded rationality in social science: Today and tomorrow. *Mind & Society*, 1(1), 25–39.

Spunt, R. P. (2015). Dual-process theories in social cognitive neuroscience. In A. Toga & M. D. Lieberman (Eds.), *Brain mapping: An encyclopedic reference* (pp. 211–215). Amsterdam: Elsevier.

Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences*, 12, 435–467.

Thagard, P. (2000). *Coherence in thought and action*. Cambridge, MA: MIT Press.

Thagard, P. (2006). *Hot thought: Mechanisms and applications of emotional cognition*. Cambridge, MA: MIT Press.

Thagard, P. (2011). Critical thinking and informal logic: Neuropsychological perspectives. *Informal Logic*, 31, 152–170.

Thagard, P. (2019a). *Brain-mind: From neurons to consciousness and creativity*. Oxford: Oxford University Press.

Thagard, P. (2019b). *Mind-society: From brains to social sciences and professions*. Oxford: Oxford University Press.

Thagard, P. (2019c). *Natural philosophy: From social brains to knowledge, reality, morality, and beauty*. Oxford: Oxford University Press.

Thagard, P., & Schröder, T. (2014). Emotions as semantic pointers: Constructive neural mechanisms. In L. F. Barrett & J. A. Russell (Eds.), *The psychological construction of emotions* (pp. 144–167). New York: Guilford.

Thagard, P., & Verbeurgt, K. (1998). Coherence as constraint satisfaction. *Cognitive Science*, 22, 1–24.

Thaler, R. H. (2015). *Misbehaving: How economics became behavioural*. New York: Norton.

van Rooij, I., Blokpoel, M., Kwisthout, J., & Wareham, T. (2018). *Cognition and intractability: A guide to classical and parameterized complexity analysis*. Cambridge: Cambridge University Press.

Wason, P. C. (1966). Reasoning. In B. M. Foss (Ed.), *New horizons in psychology* (pp. 1–43). Harmondsworth: Penguin.