JNphi

My Journey to Neurophilosophy: Paul Thagard

Paul Thagard

Abstract

Paul Thagard describes how his current work in neurophilosophy grew out of a long series of engagements with philosophy, philosophy of science, cognitive science, neural networks, and theoretical neuroscience. Each of these engagements had cumulative advantages over its predecessors. Neurophilosophy is prospering by applying insights about the workings of the brain to central problems in epistemology, metaphysics, and ethics.

Key Words: mind, brain, philosophy, neuroscience, neurophilosophy

DOI: 10.5281/zenodo.7740227

I avidly read Patricia Churchland's (1986) book *Neurophilosophy* when it came out, but was not convinced by her claims about the centrality of neuroscience to traditional philosophical concerns with knowledge and reality. At the time, I was already enthusiastic about the philosophical relevance of cognitive science, but my main interests were in psychology and artificial intelligence. I was collaborating with cognitive psychologists and developing my own computer models of high-level thinking with ample philosophical applications. Many decades later, however, I have become thoroughly convinced of the value of the neurophilosophical approach to epistemology, metaphysics, and ethics.

This article describes how my journey through philosophy and cognitive science led me toward increased involvement in neuroscience, through intermediate stages that have included philosophy of science and computational neural networks. In order to make this review more than autobiography, I also outline why my intellectual transitions make sense from the perspective of intellectual goals to understand fundamental aspects of human minds and societies. I now see neurophilosophy as central to accomplishment of philosophical ambitions towards truth and justice.

Corresponding author: Paul Thagard
Address: University of Waterloo, Philosophy Department, Waterloo, Ontario, Canada.
e-mail ✉ pthagard@uwaterloo.ca

Philosophy

I first encountered philosophy as an antidote to being a student at a Catholic high school in Saskatoon, Saskatchewan, Canada. My religious devotion had extended to being an altar boy, but by grade 9 I was having serious doubts about Catholic doctrines. My parents' educations had only gone to grade 11, but my mother was an avid reader and we made weekly trips to the Saskatoon Public Library. When I started grade 9, she put in an application for me to become a part-time employee at the library and I started spending Saturdays shelving books. One Saturday when I was 15, I came across Bertrand Russell's (1967) *Why I Am Not a Christian* which blew my mind, to use the 1960s expression. Russell convincingly demolished the standard arguments for the existence of God and demonstrated the power of a non-religious approach to knowledge and morality. Around the same time, I shelved a book in the careers section about being a professor, and got the idea that it would be cool to become a philosophy professor. I am still astonished that this ridiculously early plan worked out.

Paul Thagard in 2018.

After my family moved to Saskatoon when I was 8, my mother told me it was a good move because the city had a university. Accordingly, after high school I went to the University of Saskatchewan and became a philosophy major. I enjoyed all the courses but particularly excelled in formal logic with a grade of 100%. One of the philosophy professors who had studied at Cambridge University told me about an obscure scholarship called the Canada Scholarship at Cambridge, which was for a Canadian undergraduate to do a second BA. So after my 3-year degree in Saskatoon, I headed to Cambridge to do the Philosophy Tripos.

I only wrote much later about why philosophy is intellectually and practically superior to religion, in *The Brain and the Meaning of Life* (Thagard, 2010). The starkest difference is that philosophy can be based on reason and evidence, whereas religious faith is belief in, trust in, and devotion to gods, leaders, or texts, independent of evidence. Faith has three serious problems because religions vary enormously in what they propose, some things they propose are clearly false, and some religious practices have promoted evil actions such as the Crusades. Philosophy ought to be able to use reason and evidence to promote not only the abstractions of truth and justice but also the practical accomplishment of human wellbeing.

Philosophy of Science

At Cambridge in 1971, I had to choose between two courses of study, and picked the one on logic and philosophy of science because of my interest in logic. I had done well in high school courses in physics and chemistry because I liked the math problems, but had no general interest in science. I quickly discovered, however, that philosophy of science was an excitingly rich approach to the problems that most interested me such as the nature of knowledge. I had the good luck that the philosopher in my college, Peterhouse, was a young Canadian named Ian Hacking, who was lecturing about the history of ideas of probability (Hacking, 1975). Before arriving, he suggested I read W. V. O. Quine's (1960) *Word and Object* which took a scientific approach to philosophy of language and mind.

In the Cambridge style, I wrote weekly essays for Hacking and other instructors followed by critical responses. I also attended diverse lectures, including an illuminating series by Gerd Buchdahl (1969) concerning the relation between history of philosophy and history of science. Hacking, Quine, and Buchdahl helped me see close interconnections between science and philosophy. I attended many other courses of lectures but was unimpressed by the ones in mainstream analytic philosophy which seemed to be focused on trivial puzzles rather than deep philosophical problems. For example, one lecturer spent much of a term talking about the meaning of the sentence "A vixen is a female fox."

As reading for one of my weekly essays, Hacking assigned Noam Chomsky's (1972) *Language and Mind,* which describes language acquisition as resulting from a kind of inference that Charles Peirce had dubbed *abduction.* I started to read Peirce and encountered the modern version of abduction that Gilbert Harman (1965) called inference to the best explanation. When I went to the University of Toronto in 1973 to do graduate work, I chose this topic for my PhD thesis focusing on Peirce, Harman, and especially on examples from the history of science. My supervisor was Bas van Fraassen, but I also learned from excellent Peirce scholars, David Savan and Thomas

Goudge. This project yielded my article on the best explanation published in the *Journal of Philosophy (*Thagard 1978).

I remain convinced that historical philosophy of science is an extremely valuable approach to philosophy. For epistemology, it provides scientific theories as much more interesting examples of knowledge than the everyday, made-up examples from thought experiments that still shape mainstream analytic epistemology. For metaphysics, it provides much richer examples of explanation, truth, and falsehood than analytic story telling. Much later I adapted the term "natural philosophy" and codified the differences between scientific and non-scientific approaches to philosophy (Thagard 2012b, 2019c). I think that Quine exaggerated when he said that philosophy of science is philosophy enough, because ethics and aesthetics are not just philosophy of science, although I realized later that cognitive science is relevant to them.

Cognitive Science

Philosophy teaching jobs were scarce when I got my PhD in 1977, but I was fortunate to land a tenure-track position at the University of Michigan-Dearborn. This small branch was in a suburb of Detroit, but I was able to live in Ann Arbor where the main University of Michigan was located. By winter, 1978, I was looking for intellectual stimulation and noticed that Alvin Goldman was giving a seminar on Tuesdays, one of my non-teaching days. I was intrigued by the connections he was making between epistemology and psychology, later published in his *Epistemology and Cognition* (Goldman, 1986). But the real excitement for me came from another course, coordinated with Goldman's, that was taught on Thursdays by the social psychologist Richard Nisbett, using a manuscript of what became his book *Human Inference* (Nisbett and Ross, 1980).

This class was my first introduction to cognitive psychology which I quickly saw as relevant to my interests in scientific discovery and reasoning. In particular, psychological research on concepts as schemas seemed highly relevant to understanding the structure of scientific knowledge in ways far more precise than Thomas Kuhn's paradigms. Reading on schemas led me in February, 1978 to Marvin Minsky's (1975) paper on frames, which was my first introduction to artificial intelligence. I had no previous interest in computers, but was intrigued by the idea of computer models as a method for understanding thought.

Nisbett and I soon collaborated on several philosophy papers, and in 1980 he organized a discussion group on induction that included the computer scientist John Holland and the cognitive psychologist Keith Holyoak. We met regularly, at my apartment because the others had families, pooling our diverse understanding of learning in humans, animals, and computers. As with all

interdisciplinary work, we encountered numerous communication challenges such as terminological differences: it took us 2 meetings to figure out that we each had a different understanding of "schema". But excitement came from progress in converging on interconnected ideas, resulting in a widely-cited book (Holland, Holyoak, Nisbett, and Thagard, 1986).

In the mid-1970s, cognitive science emerged as a well-funded interdisciplinary enterprise, and The University of Michigan had a cognitive science program that I became part of, with an office in the experimental psychology building. Through extensive reading and collaborating with Nisbett and Holyoak, I became fully informed in psychology, but wanted to be able to build my own computer models. So I took graduate courses in computer science that led to an MS in 1985, enabling me to contribute to cognitive science not just as a philosopher but also as a computer modeler. In cognitive science and many other fields, computer models are valuable for formulating precise hypotheses and testing them by simulating natural phenomena. I thought I should know something about neuroscience and attended a course, but found it boring because it did not connect with my interest in high-level cognition of the kinds used by scientists.

Many years later, I published an article "Why Cognitive Science Needs Philosophy and Vice Versa" which summarizes the benefits that psychology and artificial intelligence have for philosophy (Thagard, 2009). Psychology provides a much richer account of the structure and growth of scientific knowledge than the narrow devices of analytic philosophy. Moreover, computer modeling provides a valuable tool for developing and testing complex theories about discovery and reasoning. My book *Computational Philosophy of Science* explored important applications of this cognitive-science approach to science but without any connections to neuroscience (Thagard, 1988). I am pleased that the idea of philosophers using computer models no longer seems so odd, and there is even an encyclopedia entry on computational philosophy (Grim, 2020).

Neural Networks

In 1985, I married Ziva Kunda, Nisbett's star graduate student, who became an assistant professor at Princeton University. In line with her principle that "home is where your wife is", I moved to Princeton as a visitor and then as a research scientist with external funding. Abandoning a tenured philosophy professorship was scary, but I did not see myself as abandoning philosophy, because I saw the computational, psychologically-oriented models I was developing as a better way of doing the kind of philosophy of science I valued.

1987 brought major breakthroughs which drew me into neural networks. Keith Holyoak and I had been collaborating on theorizing about analogy, trying to come up with computational models that integrated his extensive psychological experiments with my work on the role of analogy in scientific discovery. The resulting computer model was not very impressive, even to us. Fortunately, Keith was invited to review for *Science* the volumes on parallel distributed processing that developed neural network (PDP, connectionist) approaches to numerous important psychological problems such as language processing (Rumelhart and McClelland 1986). Keith recognized that the PDP approach to schema application might work for the tricky problem of analogical mapping. Just as a schema maps onto information about an object to classify it, two analogs might map onto each other by using neural networks to satisfy constraints on how they might correspond. I thought the idea was intriguing and wrote a LISP program to translate analogical comparison (e. g. Socrates is a midwife of ideas) into a process of parallel constraint satisfaction performed by simple neural networks. After a couple of weeks, I was amazed at how well the program was working, even on complicated examples. The approach resulted a series of articles and our book on analogy, *Mental Leaps* (Holyoak and Thagard, 1995).

One evening I was wondering what other problems might be amenable to similar treatment, and suddenly realized that computational constraint satisfaction offered a new approach to inference to the best explanation, where scientists have to figure out how to balance explaining the most facts while maintaining simplicity. I wrote a program ECHO that has since been applied to dozens of important examples of reasoning in science, law, and everyday life (Thagard, 1989, 1992, 1999, 2012a; Dammann, Poston, and Thagard, 2019). Philosophers since Hegel have talked about coherence as an important aspect of human thought, but have been vague about how it amounts to more than consistency. Construing coherence as maximal constraint satisfaction provides a mathematically elegant, computationally feasible, and psychologically applicable way of understanding coherence (Thagard, 2000). I applied this approach to many other issues of philosophical and psychological problems, including decision making and stereotype application.

Coherence by constraint satisfaction can be computed by traditional algorithms, but neural networks provide a biologically natural way to think about inference that is very different from traditional philosophical approaches to deductive and inductive logic. Instead of step-by-step linguistic procedures or mathematical probabilities, we can think of a problem as represented by neuron-like structures that have degrees of activation. Constraints are captured by excitatory and inhibitory connections between these structures, just as neurons are connected by excitatory and inhibitory synapses. Inference requires parallel interactions, not serial linguistic steps. I still think this is a powerful way of understanding how brains produce

powerful mental computations, so it makes valuable contributions to natural philosophy. Today it is fashionable to view the brain as a prediction engine driven by Bayesian inference, but a more comprehensive and biologically plausible approach views the brain as a coherence engine driven by constraint satisfaction.

Theoretical Neuroscience

In 1992, Ziva and I moved to tenured positions at the University of Waterloo, a fine Canadian research university. My main appointment was in the Philosophy Department, but cross-appointments to Psychology and Computer Science provided a wide range of students. I soon started a cognitive science program which attracted outstanding students from around the university.

I had long been aware that the neural-network programs I wrote to do constraint satisfaction were not biologically accurate. The artificial neurons used to represent particular elements like concepts and propositions were very different from the distributed representations among thousands of neurons that operate in the brain. I stuck with them because they enabled me to model a wide array of high-level inferences such as theory choice. I learned about wider possibilities from an amazing engineering student, Chris Elias (later Eliasmith), who became an MA student in philosophy and my research assistant. He kept insisting on the need for more biologically plausible neural networks, and found a powerful approach called holographical reduced representations developed by Tony Plate. Whereas connectionist neural networks had difficulty representing the difference between "dog bites person" and "person bites dog", Plate's methods used vectors to show how distributed neural networks could capture complex syntactic structure.

Chris followed my suggestion to do his PhD at the program in Philosophy, Psychology, and Neuroscience at Washington University in St. Louis. There he worked closely with neuroscientists in David van Essen's group, including the mathematical physicist Charles H. Anderson. They collaborated to produce a powerful and novel approach they called Neural Engineering (Eliasmith and Anderson, 2003). By this time, theoretical neuroscience (also known as computational neuroscience) was recognized as an approach to cognitive science that aimed at mathematical models of brain processes that were more biologically accurate than the connectionist, PDP models that had inspired me in the 1980s (Dayan and Abbott, 2001).

In 2004, the University of Waterloo was fortunate to hire Chris into the Philosophy Department with eventual cross appointments to engineering and computer science. He attracted strong students from multiple departments and I attended his research group meetings. Around 2009 he announced a new idea about neural representation

that I immediately recognized as a major breakthrough. He had figured out how neural representations could have the complex syntactic structure's proposed by Plate while retaining some of the sensory information that originated from the world. His new idea called "semantic pointers" seemed to me the first plausible proposal about how the brain could produce the syntactic, semantic and pragmatic complexity of human mental representations. Chris's new idea produced an article in the journal *Science*, a monumental book, and many subsequent publications (Eliasmith et al. 2012, Eliasmith 2013; Crawford, Gingerich, and Eliasmith, 2016).

With the programming assistance of Terry Stewart and Ivana Kajić, I applied Chris's semantic pointer architecture (SPA) to important problems that had long interested me: creativity, consciousness, intention, and emotion. The history of science supports the view that creativity non-mysteriously results from combining concepts that were previously unconnected, and theoretical neuroscience could now explain how this works through binding of neural representations (Thagard and Stewart, 2011). Consciousness can be explained by the binding and competition of semantic pointers (Thagard and Stewart, 2014). Intentions are neural processes that integrate representations of states of affairs, actions, and emotional evaluations (Schröder, Stewart, and Thagard, 2014). Emotions are semantic pointers that bind neural representations of situations, physiological reactions, and appraisals of goal-relevance (Thagard and Schröder, 2014; Kajic, Schröder, Stewart, and Thagard, 2019; Thagard, Larocque, and Kajic forthcoming). My post-doc Tobias Schröder showed that priming of automatic behaviors can be explained by semantic pointers (Schröder and Thagard, 2013). Chris Eliasmith's student Peter Blouw developed a comprehensive theory of concepts as semantic pointers (Blouw, Solodkin, Thagard, and Eliasmith, 2016).

My interest in emotion was prompted in 1993 by another graduate student, Allison Barnes, who proposed that empathy is a kind of analogy where you understand other people by understanding their situation based on your own emotional reactions to similar situations. I began to realize that emotion is an important part of analogy and other kinds of cognition, which led me to read Antonio Damasio path-breaking book on how rationality depends on emotion (Damasio, 1994). Damasio provided strong reason to look to neuroscience for explanations of high-level reasoning, and I recruited an engineering student to build a spiking neural network model of the phenomena he described (Wagar and Thagard, 2004). So by the 2000s I was thoroughly convinced of the central relevance of neuroscience to understanding mental processes.

I still think that Eliasmith's Semantic Pointer Architecture is the best available theory of neural processing, but it has healthy competition. Karl Friston's ideas about predictive processing and Bayesian inference have a strong following and have been extended to explain consciousness (Parr, Pezzulo, and Friston, 2022; Clark, 2015; Seth, 2021). Other researchers emphasize the computational power of deep learning and reinforcement learning (Sejnowski, 2018; Silver, Singh, Precup, and Sutton, 2021). We still lack a general theory of neural functioning that provides cognitive science with the consensual unification that biology gets from evolutionary and genetic theory and physics gets from general relativity and quantum theory. Nevertheless, the progress in theoretical and empirical neuroscience in recent decades has been astonishing, in contrast to the utter stagnation of dualist approaches to mind (Churchland, 2002, 2022).

Cognitive science has not been supplanted by theoretical neuroscience because it still gains from multiple methodologies, including the behavioral experiments of psychologists, the language theorizing of linguists, and the cross-cultural ethnologies of anthropologists. But neuroscience has progressed rapidly through ever-expanding experimental techniques such as cell recording, brain scans, and optogenetics. Understanding the empirical results requires computational models of neural mechanisms that explain how brains accomplish the full array of mental accomplishments, from reinforcement learning in insects to scientific discovery and philosophical reflection. Quine was my first exposure to naturalistic philosophy, but his work was limited by the narrow scope of behaviorist psychology. Behaviorism has long since been superseded by cognitive psychology which is increasingly integrated with experimental and theoretical neuroscience.

Neurophilosophy

Ziva's death from cancer in 2004 made me worry about the traditional philosophical question of how life can be meaningful, and I sought answers in neuroscientific understanding of human needs. The result was *The Brain and the Meaning of Life* (Thagard, 2010) which also addresses philosophical problems about knowledge, reality, and ethics.

By 2016, I had exhausted my tolerance for bureaucratic bungling and happily retired from teaching in order to write full time. I was already embarked on a three-volume *Treatise on Mind and Society* that aimed to integrate all my philosophical, psychological, and social interests (Thagard 2019a, 2019b, 2019c). During the 2010s, I collaborated with a group of political scientists interested in ideologies and social change (e.g Homer-Dixon, Milkoreit, Mock, S., Schröder, and Thagard, 2014). This work motivated me to connect more systematically the cognitive sciences with the social sciences including economics and education.

The book *Brain-Mind* applies an accessible version of the semantic pointer theory of neural processes to a full range of psychological phenomena including perception, imagery, concepts, rules, analogies, emotions, consciousness, creativity, actions, creativity, and the self. Then *Mind-Society* combines these processes with social processes of communication to apply to the full range of social sciences and processes.

The third book in the *Treatise, Natural Philosophy,* is my most extensive treatment of neurophilosophy, marking full conversion to the Churchlandian enterprise. It shows how the semantic pointer explanation of brains applies to the full range of philosophical problems concerning mind, knowledge, reality, explanation, morality, justice, meaning, beauty, and mathematics. It continues my critique of thought experiments as the primary philosophical methodology and shows the fertility of an approach that systematically connects philosophy with the sciences, especially neuroscience. I do not attempt to use science to replace philosophy, which remains of high intellectual value because of its greater generality and normativity. Writing invitations prompted follow-up articles about naturalizing logic, bounded rationality, and meaning in life (Thagard, 2021b, 2021c, 2022c). The best way to defend the value of neurophilosophy is to do it well, by showing how understanding the brain can contribute to progress on central problems in epistemology, metaphysics, ethics, and even aesthetics.

For the Waterloo cognitive science program, I developed an integrative course on intelligence in machines, humans, and other animals. My book *Bots and Beasts* worked out this comparison in great detail, based on the mechanisms for intelligence that I developed in *Brain-Mind* (Thagard, 2021a). By comparing the brains and accomplishments of humans and other animals with the structures and processes of computers, I show that current computers and non-human animals fall far short of human intelligence.

In 2016, I had a temporary problem with vertigo caused by an inner-ear disturbance, which along with a course in tai chi got me interested in the neuroscience of balance (Thagard, 2022a). Using ideas from the Semantic Pointer Architecture, I outlined a theory of how the brain combines signals from eyes, ears, and body to maintain balance, and how these mechanisms break down to produce the disturbing conscious experiences of vertigo, nausea, and falling. This extension connected my previous theory of consciousness as semantic pointer competition with alternative theories of information integration and neural broadcasting (Dehaene, 2014; Tononi, Boly, Massimini, and Koch, 2016). While thinking about balance, I kept noticing the prevalence of balance metaphors such as work-life balance, balanced diet, and balancing lives and livelihoods in the COVID-19 pandemic. The second part of my book *Balance* assesses balance metaphors as strong, weak, bogus, or toxic.

At the end of 2019, just before the pandemic hit, I had the good fortune to attend two lectures with an intriguing overlap. One was by Chris Eliasmith about the energy efficiencies of neuromorphic (brain-like) computers, and the other was by a biologist, Mary O'Connor, about the interactions between energy and information-gathering in organisms. I had not previously thought about the philosophical and psychological significance of energy requirements, but quickly developed an argument that they undermine widely accepted philosophical views about mind-body functionalism and substrate independence (Thagard, 2022b). The brain is wonderfully energy-efficient compared to electronic computers, which provides further reason to think of mind as brain rather than as abstract computation.

While working on energy, I reviewed diverse publications on the nature of information and was surprised that no rich theory had surpassed the very limited mathematical theory developed by Shannon (1948). Along with other philosophers of science, I have long maintained that many scientific theories are descriptions of mechanisms, so the question occurred to me: What mechanisms explain information? Using what I knew about how brains work, I generated a list of 8 mechanisms (Thagard, 2021d). In 2020, the news was full of complaints about misinformation concerning the pandemic, and I realized that misinformation results from breakdowns in these mechanisms, just as disease results from breakdowns in the mechanisms that promote health. This account of misinformation applies well to COVID-19, climate change, conspiracy theories, inequality, and the Russia-Ukraine war (Thagard, forthcoming). So understanding how the brain accomplishes cognition, emotion, and social interactions turns out to be relevant to the most pressing current public issues.

What next? The problem of consciousness remains pivotal to philosophical and scientific concerns about mind and reality, and I am now returning to it with some new twists. Conscious experiences fall into four basic kinds: external perceptions such as seeing and hearing, internal sensations such as pain and hunger, emotions such as happiness and sadness, and abstractions such as thinking about philosophy. I want to show that my theory of consciousness as neural representation, binding, and coherence applies to all combinations of these four kinds, in rich areas of experience such as dreams, humor, sports, music, and mental illness. I hope the result will contribute to cognitive science and neurophilosophy.

Acknowledgements

Thanks to Sultan Tarlacı for inviting this contribution.

Funding

None declared.

Conflict of interest statement

None declared.

References

Blouw P, Solodkin E, Thagard P, Eliasmith C. Concepts as semantic pointers: A framework and computational model. Cognitive Science 2016; 40: 1128-1162.

Buchdahl G. Metaphysics and the philosophy of science: The classical origins, Descartes to Kant. MIT Press; 1969.

Chomsky N. Language and mind (2 ed.). Harcourt Brace Jovanovich; 1972.

Churchland PS. Neurophilosophy: Towards a unifed understanding of the mind-brain. MIT Press; 1986.

Churchland PS. Brain-Wise: Studies in neurophilosophy. MIT Press; 2002.

Churchland PS. What is neurophilosophy and how did neurophilosoph get started? Journal of NeuroPhilosophy 2022;, 1: 1-16.

Clark A. Surfing uncertainty: Prediction, action, and the embodied mind. Oxford University Press; 2015.

Crawford E, Gingerich M, Eliasmith C. Biologically plausible, human-scale knowledge representation. Cognitive Science 2016; 40: 782-821.

Damasio AR. Descartes' error: Emotion, reason, and the human brain. G. P. Putnam's Sons; 1994.

Dammann O, Poston T, Thagard P. How do medical researchers make causal inferences? In" McCain K, Kampourakis K. eds. What is scientific knowledge? An introduction to contemporary epistemology of science. Routledge, 2019: 33-51.

Dayan P, Abbott, LF. Theoretical neuroscience: Computational and mathematical modeling of neural systems. MIT Press; 2001.

Dehaene S. Consciousness and the brain: Deciphering how the brain codes our thoughts. Viking; 2014.

Eliasmith C. How to build a brain: A neural architecture for biological cognition. Oxford University Press; 2013.

Eliasmith C, Anderson CH. Neural engineering: Computation, representation and dynamics in neurobiological systems. MIT Press; 2003.

Eliasmith C, Stewart TC, Choo X, Bekolay T, DeWolf T, Tang Y, Rasmussen D. A large-scale model of the functioning brain. Science 2012; 338: 1202-1205.

Goldman A. Epistemology and cognition. Harvard University Press; 1986.

Grim P. Computational philosophy. Stanford Encyclopedia of Philosophy. Retrieved 2020 from https://plato.stanford.edu/entries/computational-philosophy/.

Hacking I.The emergence of probability. Cambridge University Press; 1975.

Harman G. The inference to the best explanation. Philosophical Review 1965; 74: 88-95.

Holland JH, Holyoak KJ, Nisbett RE, Thagard PR. Induction: Processes of inference, learning, and discovery. MIT Press; 1986.

Holyoak KJ, Thagard P. Mental leaps: Analogy in creative thought. MIT Press; 1995.

Homer-Dixon T, Milkoreit M, Mock, SJ, Schröder T, Thagard, P. The conceptual structure of social disputes: Cognitive-affective maps as a tool for conflict analysis and resolution. SAGE Open 2014; 4: 1-20.

Kajić I, Schröder T, Stewart, TC, Thagard, P. (2019). The semantic pointer theory of emotions: Integrating physiology, appraisal, and construction. Cognitive Systems Research 2019; 58: 35-53.

Minsky M. A framework for representing knowledge. In: Winston, PH eds. The psychology of computer vision. McGraw-Hill, 1975: 211-277..

Nisbett RE, Ross L. Human inference: Strategies and shortcomings of social judgement. Prentice Hall; 1980.

Parr T, Pezzulo G, Friston KJ. Active inference: The free energy principle in mind, brain, and behaviour. MIT Press; 2022.

Quine WVO. Word and object. MIT Press; 1960.
Rumelhart DE, McClelland JL eds. Parallel distributed processing: Explorations in the microstructure of cognition. MIT Press; 1986.
Russell B. Why I am not a Christian and other essays on religion and related subjects. George Allen and Unwin; 1967.
Schröder T, Thagard, P. The affective meanings of automatic social behaviors: Three mechanisms that explain priming. Psychological Review 2013;, 120: 255-280.
Sejnowski TJ. The deep learning revolution. MIT Press; 2013.
Seth A. Being you: A new science of consciousness. Dutton; 2021.
Shannon CE. A mathematical theory of communication. The Bell System Technical Journal 1948: 27, 379-423.
Silver D, Singh S, Precup, D, Sutton RS. Reward Is enough. Artificial Intelligence 2021; 299: 103535.
Thagard P. The best explanation: Criteria for theory choice. Journal of Philosophy 1978; 75: 76-92.
Thagard P. Computational philosophy of science. MIT Press; 1988.
Thagard P. Explanatory coherence. Behavioral and Brain Sciences 1989; 12: 435-467.
Thagard P. Conceptual revolutions. Princeton University Press; 1992.
Thagard P. How scientists explain disease. Princeton University Press; 1992.
Thagard P. Coherence in thought and action. MIT Press; 2000
Thagard, P. Why cognitive science needs philosophy and vice versa. Topics in Cognitive Science 2009; 1: 237-254.
Thagard P. The brain and the meaning of life. Princeton University Press; 2010.
Thagard P. The cognitive science of science: Explanation, discovery, and conceptual change. MIT Press: 2012a.
Thagard P. Eleven dogmas of analytic phiosophy. Hot Thought. Retrieved 2012b from https://www.psychologytoday.com/ca/blog/hot-thought/201212/eleven-dogmas-analytic-philosophy
Thagard P. Brain-mind: From neurons to consciousness and creativity. Oxford University Press: 2019a.
Thagard P. Mind-society: From brains to social sciences and professions. Oxford University Press; 2019b.
Thagard P. Natural philosophy: From social brains to knowledge, reality, morality, and beauty. Oxford University Press: 2019c.
Thagard P. Bots and beasts: What makes machines, animals, and people smart? MIT Press: 2021a.
Thagard P. How rationality is bounded by the brain. In: Viale R ed. Routledge handbook of bounded rationality. Routledge; 2021b: 398-406.
Thagard P. Naturalizing logic: How knowledge of mechanisms enhances inductive inference. Philosophies 2021c; 6: 52. Retrieved from https://www.mdpi.com/2409-9287/6/2/52
Thagard P. What is misinformation? Hot Thought. Retrieved 2021d from https://www.psychologytoday.com/ca/blog/hot-thought/202107/what-is-misinformation.
Thagard P. Balance: How it works and what it means. Columbia University Press: 2022a.
Thagard P. Energy requirements undermine substrate independence and mind-body functionalism. Philosophy of Science 2022b; 89: 70-88.
Thagard P. The relevance of neuroscience to meaning in life. In: Landau I ed. *Oxford handbook of meaning in life.* Oxford University Press; 2022c: 127-144.
Thagard P. Why falsehoods fly: Fighting misinformation. Columbia University Press; forthcoming.
Thagard P, Larocque L, Kajić I. Emotional change: Neural mechanisms based on semantic pointers. Emotion 2022, Advance online publication. https://doi.org/10.1037/emo0000981.
Thagard P, Schröder T. Emotions as semantic pointers: Constructive neural mechanisms. In: Barrett LF, Russell RA eds. The psychological construction of emotions. Guilford, 2014: 144-167.

Thagard P, Stewart TC. The Aha! experience: Creativity through emergent binding in neural networks. Cognitive Science 2011; 35: 1-33.

Thagard P, Stewart, TC. Two theories of consciousness: Semantic pointer competition vs. information integration. Consciousness and Cognition 2014;, 30: 73-90.

Tononi G, Boly M, Massimini M, Koch C (2016). Integrated information theory: From consciousness to its physical substrate. Nature Reviews Neuroscience 2016;, 17: 450-461.

Wagar BM, Thagard P. Spiking Phineas Gage: A neurocomputational theory of cognitive-affective integration in decision making. Psychological Review 2004; 111: 67-79.