PhiMiSci
Philosophy and the Mind Sciences | Vol. 7 | 2026

The ontological reversal of computation and the brain

Ida Momennejad[1,*]

Book symposium on *The Brain Abstracted* by Mazviita Chirimuuta, edited by Philipp Haueis

Abstract

The Brain Abstracted (2024) critiques treating abstractions in neuroscience as complete explanations of the brain, for their oversimplification and control-orientation. Chirimuuta argues that neuroscience operates on haptic realism, where scientific knowledge arises through control-oriented experimental interaction rather than contemplative understanding of reality. She proposes distinct epistemic aims for philosophy (understanding) and science (control of nature) and suggests a disciplinary separation between the two. I argue that Chirimuuta underestimates the entanglement of philosophy and science in naturalizing abstractions, what I call ontological reversal, where abstractions gradually define the very reality they were meant to simplify. When computational models and metaphors are idealized, treated as more real than lived neural and social processes, they machinify minds and brains. Undoing the reversal of computation and brains, however, requires interdisciplinarity, rather than the book's separation of philosophy and science. This supports Chirimuuta's conceptual pluralism: treating neuroscientific abstractions not as singular truths but as evolving microcosms.

[1] Microsoft Research
[*] Primary contact: ida.momennejad@gmail.com

Momennejad, I. (2026). The ontological reversal of computation and the brain. *Philosophy and the Mind Sciences*, *7*. https://doi.org/10.33735/phimisci.2026.12308

1 Introduction

The Brain Abstracted examines whether simplifications in neuroscience and philosophy truly advance our understanding of brains and minds. The book makes three central claims. First, that while abstractions are helpful for the goals of neuroscience, they should not be read as literal descriptions of reality. That is, scientific abstractions, though pragmatic, become problematic when taken as ontological commitments about what brains are.

Second, that scientific abstraction is historically rooted in the desire to control nature, whereas philosophy aims toward understanding. Chirimuuta calls the metaphysics of neuroscience *haptic realism*, where "science constitutes its objects with a view to manipulation," and "through the interactions and iterations of experimental research" (p. 62).

Third, that these distinct epistemic goals in philosophy and science justify a disciplinary separation: Chirimuuta wants to free philosophy from the limiting constraints of abstractions that serve scientific utility, prediction, manipulation (i.e., abstractions that serve the control of nature). She believes that since philosophy does not aim to cure disease or engineer circuits, it can afford more ontological freedom in its abstractions than the sciences.

I admire the book's ambition and endorse the first claim, but challenge its division of philosophy and science. I develop the notions of the machinification of mind and ontological reversal of computation and the brain, and end by arguing that interdisciplinary practice, rather than disciplinary separation, is essential to resisting ontological reversal.

2 Understanding vs. control

> *"What is the essential object of science? It is to enlarge our influence over things"* — Bergson (1907/1944, p. 358)

Both science and philosophy translate observations into theoretical, mathematical, and empirically useful abstractions. Chirimuuta suggests that neuroscientific and mechanistic abstractions, while helpful, are rooted in science's aim of *controlling nature*. To justify this, she extends to neuroscience a famous sociology of science argument that 17th century science was the synthesis of natural philosophy's aim of understanding nature and the artisanal motif of controlling nature (Zilsel, 1942, Dear, 2005). She argues that, in contrast to science, philosophy aims at *understanding*: "since philosophy is not bound by the requirement to design conceptual tools that serve material purposes, like prediction and manipulation of physiological effects, it has more latitude in how it goes about its abstractions, and [can]... evaluate scientific abstractions by standards different from the instrumental ones of technoscience." (p. 306-307)

I am sympathetic to Chirimuuta's worry about the current direction of neuroscience becoming more control-centric, and philosophy's potential to evaluate neuroscience from outside its frameworks. As large-scale artificial neural networks and industrial AI structurally tilt research toward prediction, manipulability, and automated "discovery" (Jumper et al., 2021), we risk moving away from intelligible understanding (Evans and Duede, 2025). Like Chirimuuta, I argue that present configurations of neuroscience and neuroAI (Momennejad, 2023) intensify the very tendencies that idealize computation as the real, and brains as meaty instantiations whose biological complexities can be abstracted away.

But while Chirimuuta addresses some potential simplifications in philosophy itself (such as the duality of mind and body), she largely overlooks how philosophy's longstanding aim to control minds and societies feeds back into and shapes mechanistic abstractions in neuroscience. Moreover, while Chirimuuta correctly critiques neuroscience for treating abstractions as representations that mirror brain reality, Rorty (1979) made a similar critique about *philosophy itself*. Rorty argued that Western philosophy since Descartes has been dominated by the 'mirror of nature' metaphor, where the mind supposedly represents or reflects reality. If philosophy has long been guilty of the same representational thinking Chirimuuta criticizes in science, then the disciplinary boundary she proposes becomes harder to sustain.

This section questions Chirimuuta's philosophy-understanding vs. science-control distinction by showing that (1) the genealogy of control in contemporary neuroscience and machine learning is in fact rooted in philosophy, (2) Chirimuuta's example of personhood speaks against her distinction, and (3) Chirimuuta's claim about the lack of distality arises from an undue focus on sensorimotor neuroscience.

2.1 Controlling minds and controlling nature

Chirimuuta claims that "scientific understanding is directed toward control rather than the more contemplative purpose of understanding nature for its own sake," (p. 61) while philosophy, unconstrained by empirical or predictive demands, seeks *understanding*. She sees scientific abstractions as *instrumental*, built to manipulate, predict, and engineer (p. 62, 216, 306). By contrast, she portrays philosophical abstractions as *evaluative*, free to conceptualize beyond immediate utility (p. 306).

Chirimuuta uses Peter Dear's history of science to support her claim that scientific abstraction is rooted in the desire to control nature. Dear (2005) argues that modern science emerged from the unification of natural philosophy (contemplative understanding) with the materially directed practices of artisans and technology, such that 'natural philosophy' was "rearticulated... in the new terms of instrumentality" (Dear, 2005, p. 397). However, Dear's historical argument does not trace neuroscience, nor the aim to control *people* or *minds*. In extending Dear's argument to neuroscience, associating philosophy with contemplative understanding and neuroscientific abstraction with practical control, Chirimuuta ignores philosophy's aims to control *people* as analogous to science's aim to control nature. This poses the question: does she consider brains merely a part of 'nature' or also a part of 'people'? This is essential to her argument, given philosophy's ancient aim to control people.

Far from being immune to the logic of control, philosophy often establishes its norms. Philosophy has long served the project of world-shaping and control. For millennia, philosophy has engaged in aligning societies, institutions, and cosmologies with ontologies of human vs. the divine, the state, and the cosmos. I argue that philosophy's far more ambitious goals of control have played a formative, perhaps dominant, role in the genealogy of experimental control in neuroscience.

While *controlling nature* is a central aim in engineering (e.g., building practical things, which requires causal, physical, and mechanistic understanding) and medicine (e.g., controlling disease, which requires biological understanding); *controlling minds*, people and their behavior, is a central aim in philosophy (e.g., Hobbesian sovereign, Bentham's Panopticon, which require understanding people with utility and manipulation). While the control of nature and the control of minds are entangled in the study of the brain, Chirimuuta uncritically adopts Dear's genealogy of control in physics and chemistry, entirely sidestepping the aim of controlling minds and behavior. This erases the foundational role of philosophy in neuroscience, given the ascription of mental faculties to brain regions goes back to Aristotle.

Consider the notions of utility maximization and optimization. While shaping the core of biological learning (Thorndike, 1927), evolution (Frank, 2009), and artificial intelligence (Turing, 1950), these simplifications go back to Bentham's utilitarianism (Bentham, 1789), following Hobbes' view of human behavior as governed by rational self-interest under sovereign order (Hobbes, 1651). Utilitarianism is at the heart of late 19th and early 20th century behaviorism (Watson, 1913, Skinner, 1938), which in turn influenced neuroscience, machine learning, and reinforcement learning (Sutton, 1988, Sutton and Barto, 1998), as well as Bayesian (Griffiths et al., 2008) and evolutionary models (Frank, 2009). Notably, these philosophies made strong assumptions about human nature, what humans ought to do, and how humans ought to be governed, which shaped today's modern political and economical systems.

This genealogy creates a self-reinforcing cycle: utilitarian philosophy shapes behavioral science, neuroscience, and machine learning, which in turn provide tools to engineer more utilitarian behavior, which is taken as validating the original philosophical assumptions.

I would even argue that when science serves control, philosophy supplies the blueprints. To influence people and societies, science must often

adopt or be framed within a broader philosophical worldview, often rooted in domination (Horkheimer and Adorno 1947/2002, see Chirimuuta's footnote 13, chapter 8.2). Scientists may disavow metaphysics, but the metaphysics often speaks through them nonetheless: latent in the language, methods, and goals of inquiry. Science's unspoken metaphysical commitments, like linear causality or individualism, are not neutral. They come from somewhere, often from philosophy.

2.2 Personhood in cognitive neuroscience

To justify the divergence between neuroscience and philosophy, Chirimuuta suggests that scientific abstractions only address the subpersonal, and cannot address personhood:

> The philosopher, but not the scientist, has available a concept of personhood that takes persons to be embedded in an expansive network of circumstances and does not abstract away from subjectivity, intersubjectivity, and the normativity that is an omnipresent aspect of this form of existence. (p. 307)

I think this is not true of all philosophy nor all neuroscience. Many (analytic) philosophers make simplifications that do not consider interconnected notions like personhood and intersubjectivity. Moreover, if personhood is a key example, then the book should not have merely focused on sensorimotor neuroscience but on areas where person-level phenomena play a key role (e.g., autobiographical memory, planning, decision making, social cognition etc.).

Much cognitive computational neuroscience of memory aims to identify how memory is structured in the brain to simultaneously capture the past and enable predicting, reasoning about, and planning the future. Modeling and neuroscientific work finds that memory is structured as multiscale predictive representations (Momennejad et al., 2017, 2018, 2021, 2020, 2025). Empirical and computational accounts of how distal memories affect our present and future can help explain continuity in personhood. Research on collective cognition addresses intersubjectivity: groups of interacting humans communicate their memories, and the structure of these interactions aligns the member's forgetting, recall, and neural "representations" (Coman et al., 2016, Momennejad et al., 2019, Brunec and Momennejad, 2021). The field also considers autobiographical memory (Conway and Pleydell-Pearce, 2000, Cabeza and St. Jacques, 2007), and how memory interacts with environmental, developmental, emotional, and social circumstances during both encoding and retrieval (McEwen and Morrison, 2013, Nadel and Hardt, 2011).

Considering these examples, it is difficult to accept Chirimuuta's claim that only philosophy can capture the concepts of personhood and intersubjectivity as "embedded in an expansive network of circumstances" (p. 307). Moreover, the cognitive neuroscience of memory and decision-making, executive function, and collective cognition explicitly draw on rich philosophical traditions (Bratman, 1999), further undermining a neat philosophy-science distinction.

2.3 Distality and causal explanation in cognitive neuroscience

On a related note, Chirimuuta highlights Mayr's distinction between proximate and ultimate causal explanations in biology, taking the latter as legitimate departures from reductive proximality (p. 154-157). She also contrasts linear stimulus-response causality with circular causality involving reciprocal organism-environment interactions (p. 84-86). She then argues that cognitive neuroscience must focus on distal relationships between neural activity and environmental states, rather than tracing contiguous causal chains through intermediary mechanisms (Chirimuuta, 2024, p. 152-168).

This framing is compelling, except that she relies primarily on sensorimotor neuroscience: retinal ganglion cells tracking visual features (p. 158) and neurons responding to faces (p. 160-161), while ignoring areas of neuroscience that already endorse distality. For instance, in her discussion of face-selective neurons, she does not consider studies that explain face-selective firing distally, e.g., neurodevelopmental research on the role of exposure to

faces in early years in the formation of the face domain (Arcaro et al., 2017). Moreover, truly distal phenomena receive minimal attention: memory systems connecting past experiences to future predictions across extended timescales, long-term planning, autobiographical memory spanning lifetimes, cognitive maps across space and time (Tolman, 1948, Momennejad, 2020), and collective cognition (Momennejad, 2021).

In short, many subfields of cognitive neuroscience explicitly engage with distality, far beyond sensorimotor domains. This mirrors the concern with her personhood argument: claiming cognitive neuroscience cannot address certain phenomena while focusing on subfields least likely to address them. In this, the book commits what it warns against: it over-simplifies neuroscience, searching for personhood and distality under the lamp of receptive fields.

3 Ontological reversal

> *"The apparatuses of observation are productive—they do not just measure but help constitute the phenomenon."* — Barad (2007)

Section II touched on the descriptive-normative feedback loops between philosophy and the neurosciences. Here, in agreement with Chirimuuta, I consider how abstractions and metaphors, philosophical or scientific, exert control over reality. I propose that interdisciplinarity with true mastery over multiple disciplines, rather than disciplinary autonomy, is necessary for a way out.

3.1 Naturalized abstractions, machinified minds

Metaphors do not merely describe minds but can actively reshape them (Dreyfus, 1972, Dupuy, 2000, Foucault, 2003 and 2008, Varela et al., 1991). When humans are incentivized to behave in machine-like ways, not only do the underlying philosophies appear vindicated, but minds begin to mirror machines. This idea closely resonates with two slogans coined by Ian Hacking: 'Making up people' to refer to "the ways in which a new scientific classification may bring into being a new kind of person" (Hacking, 2007, p. 286), and 'The looping effect' to refer to "the way in which a classification may interact with the people classified".

This naturalization of machine metaphors is what I call the *machinification* of minds, when tools and models meant for understanding the mind come to reshape the mind, and historically contingent categories and worldviews are taken for granted as descriptions of reality itself (Haslanger, 2012, 2017). This echoes what Dahlin, following Husserl's critique of the mathematization of experience (Husserl, 1970, Harvey, 1989), calls *ontological reversal*, where "abstract mathematical models are seen as more real than the concrete, lived experience …from which they have been abstracted… what actually is secondary, ontologically speaking, becomes primary." (Dahlin et al., 2009)

In contemporary neuroscience, dominant computational abstractions become so widely accepted that studying them is deemed identical to studying the brain (Choudhury and Slaby, 2012). Dominant models are treated as inevitable descriptions of the brain's nature, shaping both the questions scientists ask and the interpretations they deem acceptable. Chirimuuta points out that retinotopic maps (simplified, idealized representations of how visual space is organized in the brain), taxonomies of cell types, and artificial neural networks gained disciplinary authority not only because of their predictive success, but also through institutional, educational, and methodological reinforcement. Similarly, optimization-based models of intelligence (Silver et al., 2021), originally introduced as useful simplifications, gradually become hegemonic: invisible as metaphors and instead assumed as natural law. No longer regarded as modeling choices, computation and optimization simply become how brains work (see Choudhury and Slaby, 2012).

The consequence is a subtle gradual inversion: computational abstractions, originally meant as simplified tools for making science tractable, acquire a normative authority over the real. First, they equate the study of the model with the study of the brain itself, and later, become the normative model of how minds and brains *should* function. Thus, hegemonic abstractions perpetuate normative assumptions about cognition, intelligence, and

how minds and brains work.

Human minds and brains, after ongoing ontological reversal of computationalism, are expected to be understood with simple laws of efficiently optimizing rewards: like machines. These expectations then feed back into institutional, educational, and social norms. Over time, epistemic models become moral-imperatives and identity scripts, creating systems that reward conformity to computational norms while punishing deviation.

3.2 Undoing ontological reversal

The analysis presented here reveals why interdisciplinary practice matters for studying minds and brains. Cultivating simultaneous mastery across philosophy and neuroscience, each offering different ontologies, methods, and constraints, takes more time and effort than disciplinary separation. However, it affords the interdisciplinary researcher the unique ability to evaluate each field from both within and outside itself. Rather than choosing one discipline's perspective as authoritative or following hegemonic abstractions, mastery across multiple disciplines can reveal what each framework's abstractions capture and what they obscure or entirely miss. Such a practice is not mere perspectival sampling, it requires a sustained, sometimes alternating, sometimes integrating, engagement that resists the naturalization of any single framework.

Rather than disciplinary separation, I would argue for an interdisciplinary position that seeks mastery of multiple scientific and philosophical fields. I am not suggesting that all scientists or philosophers should engage in an arduous commitment to all fields. I am, however, proposing that undoing ontological reversals, where abstractions replace the real and serve as its ideal, calls for a sizable interdisciplinary practice.

4 Conclusion: Living microcosms of understanding

To conclude, the goal of understanding brains cannot be achieved with a separation of philosophy from science, e.g., based on their supposed genealogical differences in seeking control versus understanding. Both disciplines share vulnerability to naturalizing their abstractions or what I've called ontological reversal, hence muddying control and understanding, and both share responsibility for maintaining the visibility of their conceptual choices. The study of minds and brains sits at the intersection of philosophy's historical project of governing subjectivity and science's project of manipulating nature, making it especially susceptible to ontological reversal. By revealing how neuroscientific abstractions become hegemonic, and therefore invisible, we highlight the importance of maintaining conceptual pluralism to honor the brain's perpetual flux.

Takashi Kuribayashi's installation (*For Trees*, 2015) offers a striking closing image here. Kuribayashi placed cut-up sections of tree trunk in glass boxes, creating what seemed like artificial, lifeless separations. Yet over time, these sections decayed and gave birth to new organisms and ecosystems, each glass box becoming a tiny world unto itself. The installation reminds us that even imposed separations can generate unexpected forms of life. Perhaps we need precisely the simultaneous consideration of such "living microcosms" of abstractions as multiple, evolving, and visible choices to capture the processual nature of brains, behavior, and environment across scales.

Undoing ontological reversal requires deep interdisciplinary practice. Repeatedly shifting between philosophy and neuroscience's abstractions, like a Necker cube, can keep any single abstraction from calcifying into "nature", making room for renegotiating what minds and brains are taken to be. Just as Kuribayashi's boxed tree trunks unexpectedly sprouted new life, our abstractions, kept alive as plural choices, visible as conceptual tools rather than hegemonic truths, can undo the ontological reversal, renaturalizing brains in flux.

References

Arcaro, M. J., Schade, P. F., Vincent, J. L., Ponce, C. R., & Livingstone, M. S. (2017). Seeing faces is necessary for face-domain formation. *Nature Neuroscience*, *20*(10), 1404–1412. https://doi.org/10.1038/nn.4635

Barad, K. (2007). *Meeting the universe halfway: Quantum physics and the entanglement of matter and meaning*. Duke University Press.

Bentham, J. (1789). *An introduction to the principles of morals and legislation*. T. Payne.

Bergson, H. (1944). *Creative Evolution* (A. Mitchell, Trans.) [Original work published 1907]. Random House.

Bratman, M. E. (1999). *Faces of intention: Selected essays on intention and agency*. Cambridge University Press.

Brunec, I. K., & Momennejad, I. (2021). Predictive representations in hippocampal and prefrontal hierarchies. *Journal of Neuroscience*, *41*(47), 9912–9924. https://doi.org/10.1523/JNEUROSCI.1327-21.2021

Cabeza, R., & St. Jacques, P. (2007). Functional neuroimaging of autobiographical memory. *Trends in Cognitive Sciences*, *11*(5), 219–227. https://doi.org/10.1016/j.tics.2007.02.005

Chirimuuta, M. (2024). *The brain abstracted*. MIT Press.

Choudhury, S., & Slaby, J. (2012). *Critical neuroscience: A handbook of the social and cultural contexts of neuroscience*. John Wiley & Sons.

Coman, A., Momennejad, I., Drach, R. D., & Geana, A. (2016). Mnemonic convergence in social networks: The emergent properties of cognition at a collective level. *Proceedings of the National Academy of Sciences*, *113*(29), 8171–8176. https://doi.org/10.1073/pnas.1525569113

Conway, M. A., & Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self-memory system. *Psychological Review*, *107*(2), 261–288. https://doi.org/10.1037/0033-295x.107.2.261

Dahlin, B., Østergaard, E., & Hugo, A. (2009). An argument for reversing the bases of science education: A phenomenological alternative to cognitionism. *Nordic Studies in Science Education*, *5*(1), 17–29. https://doi.org/10.5617/nordina.839

Dear, P. (2005). What is the history of science the history of? Early modern roots of the ideology of modern science. *Isis*, *96*(3), 390–406. https://doi.org/10.1086/447747

Dreyfus, H. L. (1972). *What computers can't do: A critique of artificial reason*. Harper & Row.

Dupuy, J.-P. (2000). *The mechanization of the mind: On the origins of cognitive science*. Princeton University Press.

Evans, J., & Duede, E. (2025). After science. *Science*, *390*(6655), eaec7650. https://doi.org/10.1126/science.aec7650

Foucault, M. (2003). *Society must be defended: Lectures at the Collège de France, 1975–76* (M. Bertani & A. Fontana, Eds.; D. Macey, Trans.). Picador.

Foucault, M. (2008). *The birth of biopolitics: Lectures at the Collège de France, 1978–79* (M. Senellart, Ed.; G. Burchell, Trans.). Palgrave Macmillan.

Frank, S. A. (2009). Natural selection maximizes Fisher information. *Journal of Evolutionary Biology*, *22*(2), 231–244. https://doi.org/10.1111/j.1420-9101.2008.01647.x

Griffiths, T. L., Kemp, C., & Tenenbaum, J. B. (2008). Bayesian models of cognition. In R. Sun (Ed.), *Cambridge Handbook of Computational Psychology* (pp. 59–100). Cambridge University Press.

Hacking, I. (2007). Kinds of people: Moving targets. *Proceedings of the British Academy*, *151*, 285–318. https://doi.org/10.5871/bacad/9780197264249.003.0010

Harvey, C. W. (1989). *Husserl's phenomenology and the foundations of natural science*. Ohio University Press.

Haslanger, S. (2012). *Resisting reality: Social construction and social critique*. Oxford University Press.

Haslanger, S. (2017). Critical theory and practice. In I. J. Kidd, J. Medina, & G. J. Pohlhaus (Eds.), *The Routledge Handbook of Epistemic Injustice* (pp. 455–464). Routledge.

Hobbes, T. (1651). *Leviathan*. Andrew Crooke.

Horkheimer, M., & Adorno, T. W. (2002). *Dialectic of enlightenment: Philosophical fragments* (E. Jephcott, Trans.) [Original work published 1947]. Stanford University Press.

Husserl, E. (1970). *The crisis of the european sciences and transcendental phenomenology* (D. Carr, Trans.) [Original work published 1936]. Northwestern University Press.

Jumper, J., Evans, R., Pritzel, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2

Kuribayashi, T. (2015, December). For trees, installation. https://www.takashikuribayashi.com/works?lightbox=image_2185

McEwen, B. S., & Morrison, J. H. (2013). The brain on stress: Vulnerability and plasticity of the prefrontal cortex over the life course. *Neuron*, *79*(1), 16–29. https://doi.org/10.1016/j.neuron.2013.06.028

Momennejad, I. (2020). Learning structures: Predictive representations, replay, and generalization. *Current Opinion in Behavioral Sciences*, *32*, 155–163. https://doi.org/10.1016/j.cobeha.2020.02.017

Momennejad, I. (2021). Collective minds: Social network topology shapes collective cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *377*(1843), 20200315. https://doi.org/10.1098/rstb.2020.0315

Momennejad, I. (2023). A rubric for human-like agents and neuroAI. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *378*(1876), 20210446. https://doi.org/10.1098/rstb.2021.0446

Momennejad, I. (2025). Memory and planning in brains and machines: Multiscale predictive representations. In L. Nadel & S. Aronovitz (Eds.), *Space, Time, and Memory*. Oxford University Press. https://doi.org/10.48550/arXiv.2401.09491

Momennejad, I., Duker, A., & Coman, A. (2019). Bridge ties bind collective memories. *Nature Communications*, *10*, 4642. https://doi.org/10.1038/s41467-019-12577-9

Momennejad, I., Otto, A. R., Daw, N. D., & Norman, K. A. (2018). Offline replay supports planning in human reinforcement learning. *eLife*, *7*, e32548. https://doi.org/10.7554/eLife.32548

Momennejad, I., Russek, E. M., Cheong, J. H., Botvinick, M. M., Daw, N. D., & Gershman, S. J. (2017). The successor representation in human reinforcement learning: Evidence from retrospective revaluation. *Nature Human Behaviour*, *1*, 680–689. https://doi.org/10.1038/s41562-017-0180-8

Nadel, L., & Hardt, O. (2011). Update on memory systems and processes. *Neuropsychopharmacology*, *36*(1), 251–273. https://doi.org/10.1038/npp.2010.169

Rorty, R. (1979). *Philosophy and the mirror of nature*. Princeton University Press.

Silver, D., Singh, S. P., Precup, D., & Sutton, R. S. (2021). Reward is enough. *Artificial Intelligence*, *299*, 103535. https://doi.org/10.1016/j.artint.2021.103535

Skinner, B. F. (1938). *The behavior of organisms: An experimental analysis*. Appleton-Century-Crofts.

Sutton, R. S. (1988). Learning to predict by the methods of temporal differences. *Machine Learning*, *3*(1), 9–44. https://doi.org/10.1007/BF00115009

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. MIT Press.

Thorndike, E. L. (1927). The law of effect. *American Journal of Psychology*, *39*, 212–222.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, *55*(4), 189–208.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, *49*, 433–460.

Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. MIT Press.

Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review*, *20*(2), 158–177. https://doi.org/10.1037/h0074428

Zilsel, E. (1942). The sociological roots of science. *American Journal of Sociology*, *47*(4), 544–562.