Jeffrey Kane. *The Emergence of Mind: Where Technology Ends and We Begin*. Cham, Switzerland: Palgrave Macmillan, 2024. 170pp. ISBN-10: 3031468341; ISBN-13: 9783031468346.

**Review by Damon Kutzin, University of Cambridge (dtk23@cam.ac.uk)**

Jeffrey Kane's new book *The Emergence of Mind: Where Technology Ends and We Begin* has arrived at seemingly the perfect time. Discussion surrounding artificial intelligence (AI) has reached a fever pitch in our culture, prompted by the undeniably impressive achievements of LLMs like ChatGPT and Claude. Kane describes the situation aptly: "I write in the belief that there are many individuals who, like me, find the sheer power and sophistication of Generative Artificial Intelligence both awe inspiring and disturbing. Daily articles flood the internet attempting to demystify its processing systems. Others, just as often, raise the alarm that this new technology will, if left unchecked, destroy personal privacy, if not humanity itself" (ix). It has also led to a crisis of self-understanding. Never has an AI system passed a Turing test so effectively, raising anew the question Turing posed in 1950: can machines think? And if they can so convincingly imitate natural language, mathematical proof, and literary creation, what real difference is there with how *we* think?

If there is a list of ideal philosophers who can help us think through this question, Michael Polanyi must surely be on it. Whether you agree with him or not, his theory of tacit knowing provides crucial challenges to those thinking through this question and its implications. However, AI has changed significantly from his time, raising the possibility that old philosophical critiques about AI no longer apply to the current power of machine learning. Kane's book is thus doubly welcome: it updates the arguments against a computational theory of mind to address the most recent developments in AI, and it does so through a significant, though not exclusive, engagement with Polanyi.

In the first chapter, Kane indicates that the book's purpose is to challenge the idea that generative AI is "the *embodiment* of the human mind" and to offer an alternative in its place "where life and consciousness are central to the creation of human thought" (7). Kane also promises to achieve this in less than 200 pages, taking us on a whirlwind tour of generative AI, philosophies of mind, biological origins of consciousness, and reductionist theories of science. This seems a tall order, and Kane acknowledges that he is not presenting an exhaustive study of these fields. Rather he is sacrificing precision to present a synoptic vision—not a "final statement" but an "invitation to dialogue" (8). While he is right that narrow disciplinary concerns can make us lose the forest for the

 © 2026 by the Polanyi Society

trees, the risk of a book like this is that it passes over the trees too quickly. Kane's alternative vision could seem well supported by the sheer number of argumentative pillars he produces in its favor, but if each pillar is shaky, the whole vision could collapse.

Kane's book can be divided into three parts. The first part (chapters 2–4) articulates the conceptual history and limitations of a computational theory of the mind. The second part (chapters 5–6) discuss the prospects of a reductionist science in general, focusing on physics and biology. He argues that even in physics there are emergent complex systems which cannot be reduced to those laws at lower levels, and in biology the emergence of self-organizing systems (living organisms) hammers home the implausibility of reductionism. The third and last part (chapters 7–8) presents his alternative conception of mind. I'll go over each part in this review.

The first part (chapters 2–4) begins by articulating the theory of mind he seeks to reject. It is less the theory of any particular philosopher or scientist than a general conceptual framework present in many research programs and the general zeitgeist. It consists, first, of reductive physicalism, that "all that exists and all that can exist in the physical universe owes its existence and attributes to the properties of the most elementary of all particles of matter. There are no exceptions, not even for the human mind" (11). In this framework, complex neural processing is sufficient to explain conscious mental states, and therefore there is no principled reason why a non-biological information processor couldn't perform the same functions.

The framework also blends this physicalism with a rigidly empirical methodology: that we can only understand the mind through its observable effects. This has roots in behaviorism, which tried to understand human and animal thought with reference to observable behavior alone. We can see echoes of behaviorism in Turing's claim that if computers succeed at the imitation game, there is no further question to be asked about whether they think. For contemporary theories, it involves restricting their explanatory framework to input-output relations or predictive power; there is no reckoning with subjectivity or meaning.

Contemporary computational theories of the mind tend to presuppose this "empirical-reductionist" framework in their research. They reason generally as follows: we cannot observe the inner workings of the mind, but we can observe its outputs, say, in terms of language. We can find statistical patterns in these outputs and model them in algorithms which produce similar patterns. In very general terms, this is how LLMs work. They absorb billions, sometimes trillions, of statistical patterns in the language on the internet, creating a probability web

 © 2026 by the Polanyi Society

of word associations. This web sometimes evolves into a "world model" because, over time, these associations approximate causal relationships that hold in the real world. It infers facts about the world from regularities in text, like that Paris is the capital of France or water extinguishes fire, and can apply them successfully in novel contexts. Although not many researchers would claim that current LLMs are thinking the way humans do, many believe that as the models scale (absorb more data, have more parameters, more training, etc.), then more human-like thinking will emerge.

The success of these models in studying the structures and processes of human thought and language, combined with a physicalist assumption that mind is just a kind of neural-information processing, leads to an irresistible conclusion: "Not only did it seem that the principles of human mind and principles of computational machines were one and the same, but the latter became the operative model for both" (45). Kane compellingly describes this state of affairs as "the inversion of the metaphor" of the mind as a computer: "the metaphor of the mind as a computational system inverts the product and the prototype[…]computation as the prototype for the mind mistakes a limited product of its [the mind's] activity to define the principles that guide the effort" (47). While undoubtedly useful, proponents of these theories tend to make philosophical claims beyond what is justified by the research.

The second part of the book (chapters 5–6) concerns the implausibility of reductionism and the emergence of life and mind. In these two chapters, Kane primarily appeals to scientific work to undermine reductive physicalism. He joins the ranks of many philosophers who seek to understand life or mind from a scientific point of view without thereby reducing either to the physical. To do so, he will have to address in some form what philosophers call "the causal closure of the physical": that all physical effects have only physical causes. If mental states or animate forces were to have physical effects (like the intentional movement of my arm or an acorn growing into a tree), they must ultimately be identifiable with physical causes.

He starts by arguing that, even in physics, there are complex systems which, although they do not modify underlying laws, still express new and emergent organizational principles which exercise downward causation on their component parts. Kane draws on the work of theoretical physicists to argue that a popular view of physics is false: the universe is not intelligible in terms of individual particles governed by strict laws but "is filled with complex systems that include dynamic interactions in and between the constellations of matter and energy. These relational aspects of the systems have causal power, and are at the heart of the evolution of the physical universe…and everything within it" (79).

 © 2026 by the Polanyi Society

Ilya Prigogine's theory of dissipative systems is highlighted to illustrate how complex systems from physics to biology generate their own organizational structure over and against the physical laws of their surroundings (83). Such systems cannot be explained with reference to their lower-level parts alone.

Kane ends up articulating a position similar to Polanyi's discussion of "boundary conditions": lower levels of reality (like physics and chemistry) provide constraints on higher levels (biological systems, minds, machines) without determining their operation, while the higher levels "harness" the elements of the lower levels for their purposes. But like Polanyi, it is ambiguous what kind of emergence this implies.

Take the following quote (97; emphasis added):

> Living organisms introduce *a new generative force* into the physical world that allows them to act selectively with reference to their own unique interests. That organizational force simply does not exist elsewhere. The moon does not orbit around the earth because it anticipates a benefit in doing so. Electrons do not wheel around atomic nuclei because they have their own agendas.

Language of "a new generative force" suggests that Kane subscribes to *strong emergence*. In other words, that life operates by principles of a different ontological kind to physical laws, which has its own causal power. But in other sections, he seems to take great care to show how emergent properties are continuous with the levels on which they supervene (96; emphasis added):

> The emergent principles organizing and operating in complex systems *do not introduce new laws of physics or new laws of chemistry*. Rather, *they constrain* the range in which the laws express themselves; *they limit* the possibilities for action; they create order where shear random probabilities otherwise would hold sway.

In this passage he describes the causal force of higher-level principles in much weaker terms, as "constraining the range" of lower-level forces. This ambiguity creates problems because it obscures the range of alternative conceptions of life and mind we might have. If we accept that complex systems operate according to principles that are non-reducible to those of a lower level, we do not have to accept that such higher-level principles are novel causal forces or properties. Rather they could just be *epistemically* emergent: our theories about what's going on at higher levels cannot be reduced to a lower level now, but *in principle* they

 © 2026 by the Polanyi Society

could and *ontologically* they are still physical. Polanyi scholar Jon Fennell reads Polanyi in this latter way, for instance, despite all of Polanyi's poetic waxing on emergence (2017).

Kane's rhetoric similarly suggests he subscribes to strong emergence, especially when it comes to biological beings. But, in my opinion, the evidence he brings forth is compatible with a variety of "emergences." Regardless, this point of criticism does not affect his central aim of undermining computational theories of the mind. Although it may weaken the motivation for rejecting them *a priori*, the scientific examples and phenomena Kane produces still pose a significant empirical challenge for such theories to overcome.

In part 3, the final two chapters of the book, Kane completes the picture of mind as rooted in many levels of being: our bodies, evolutionary history, cultural history, lived experience, linguistic systems, and social settings. He discusses how these levels are operative in all knowing, which vividly contrasts with the assumptions of computational theories discussed in the first chapters. In particular, his discussion of how children learn to understand the world around them is instructive. There is a superficial similarity between children and AI in that to understand the principles behind observable behaviors, they imitate those behaviors. Generative AI, through a process called unsupervised learning, also imitates the patterns it sees through vast amounts of raw data and generates novel but reliable outputs from them. But for children, "that imitation, however, is not a parroting but a projection of their subjective experience into the actions they see in others with whom they share consciousness" (147-48). This kind of learning is importantly *transformative* in a way that it isn't for AI: "The schema acquired through imitation and imagination not only inform the child's actions but become cognitively, emotionally, and socially transformative[…] All learning shares the same basic structure of internalizing the principles that govern the world we encounter" (149).

Genuine learning is the result of what Polanyi calls *indwelling*, a combination of internalization and transformation which has no equivalents in computational theories of the mind. And how could it? Our thinking is rooted in our bodies and conceptual schema that "includes everything that rises from our genetics going back over time to the beginnings of life, all that we have gained through the shared awareness that has shaped culture, and personal-lived experience" (150). Given all of the above, the reader may wonder how the hypothesis that the mind is machine was ever so plausible. In Noam Chomsky's words, asking whether machines think seems a bit like asking whether submarines swim; it depends how we choose to extend the meaning of these words.

 © 2026 by the Polanyi Society

Kane ends the book by suggesting that our readiness to metaphorically extend the notion of thinking reflects a crisis in our self-understanding. We have "learned not to think" because so much of our education is now rote rule-application, "and computationalists are correct in one critical respect: thinking independent of subjective inner activity is information processing" (151). The best illustration of this is the tedious education in mathematics most of us had to endure in school, which only reinforces the idea that learning is storing information and mechanically applying rules. The contrast between our typical math education and genuine, imaginative mathematical thinking is well illustrated in the anecdote Kane relates about Carl Friedrich Gauss's boyhood talent (152):

> His teacher asked his class to add all the numbers from 1 to 100, probably hoping to keep his pupils quiet for a half-hour. But little Gauss immediately raised his slate with the result. He had rapidly perceived the symmetry of the problem. By "mentally folding" the number line, he could group 100 with 1, 99 with 2, 98 with 3, and so on. Hence the sum was reduced to 50 pairs, each totaling 101, for a grand total of 5,050.

The role of imagination and intuition in genuine thinking is clearly seen in this simple example, never mind the more complex scientific discoveries of Einstein and McClintock that Kane also discusses in depth. Kane worries that, as a society, we are forgetting their role, and so "the language we use to describe the human mind and ourselves as human beings will slowly evolve to fit within the parameters set by technology as our cognitive touchstone" (153).

Overall, Kane's book succeeds in its mission of complicating computational theories of mind. The synoptic vision he presents has the added advantage of being accessible to a wide audience, making it an ideal introduction for those seeking to navigate the discourse around AI and consciousness. By presenting in an accessible form the science of complex systems, evolutionary emergence, and developmental psychology, Kane complicates a simplistic picture of science as reducing the human mind to computation. In doing so, he also challenges a simplistic but popular view of what it is to be a naturalist. The interpretation of our best science does not imply a reductive physicalism; rather, the emergence of life and mind from dead matter passes through multiple levels of organization (physical, chemical, biological, social), each of which cannot be understood solely through the laws and principles of a lower level.

 © 2026 by the Polanyi Society

**Works Cited**

Fennell, Jon. “Polanyi, Emergence, and Common Sense: Reflections on the Paksi Essay.” *Appraisal* 11, no. 3 (2017): 47–51.

 © 2026 by the Polanyi Society