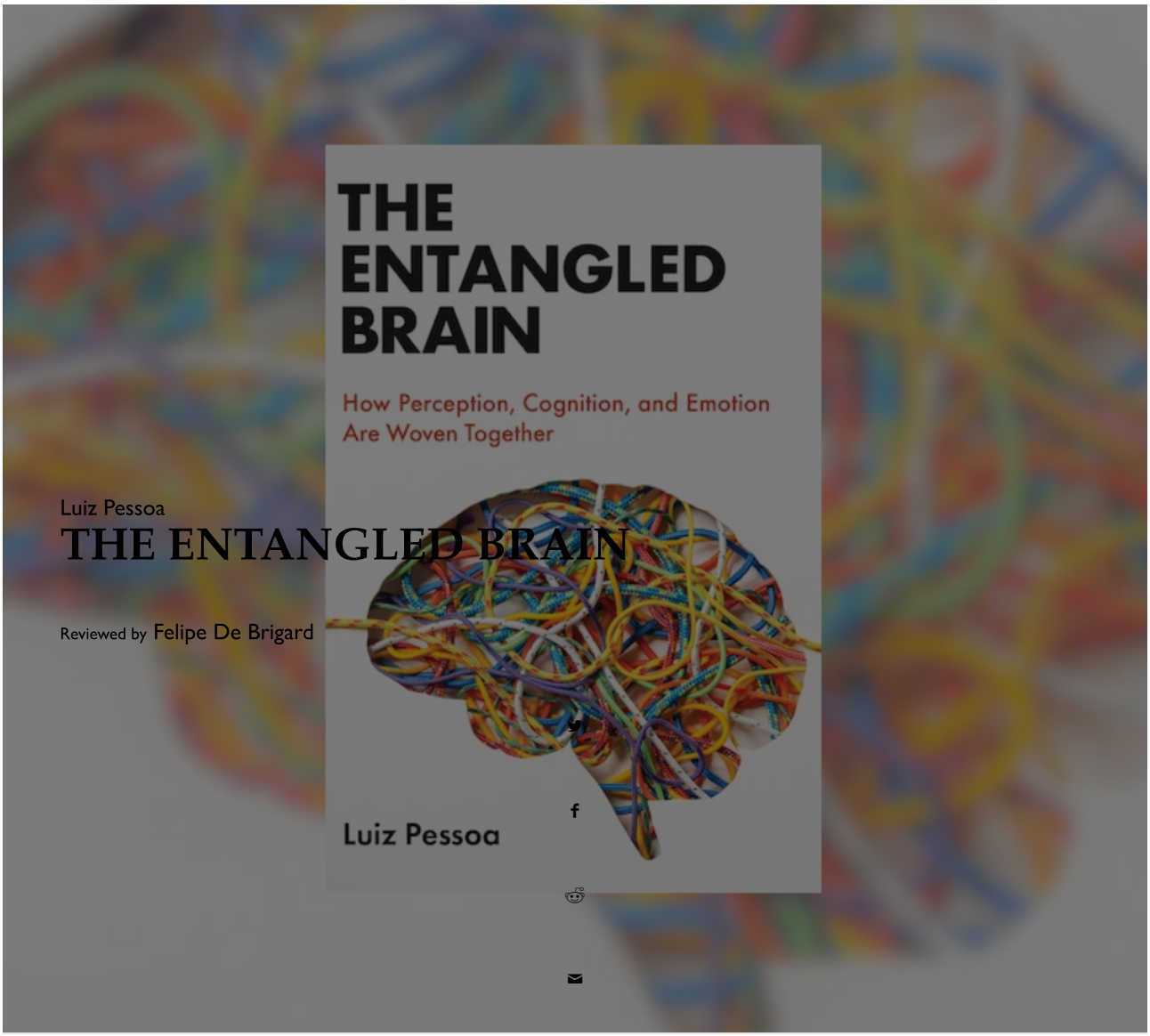

THE ENTANGLED BRAIN

Luiz Pessoa

Reviewed by
Felipe De Brigard

The Entangled Brain: How Perception, Cognition, and Emotion Are Woven Together
Luiz Pessoa
Cambridge, MA: MIT Press, 2022, £xxxx
ISBN 9780262544603

Cite as:
De Brigard, F. [2024]: 'Luiz Pessoa's *The Entangled Brain*', *BJPS Review of Books*, **2024**

The theoretical physicist Michio Kaku ([2014]) once stated that the brain is 'the most complicated object in the known universe'. For decades, neuroscientists have been trying to disentangle the brain's complexity in order to understand how it can support our behaviours and mental life. In his latest book, Luiz Pessoa wants us instead to embrace the entanglement of this intricate organ, not as a way to give up on our quest to understand its workings, but as a change in strategy to better comprehend its complexity.

To begin to motivate the claim that the brain is not decomposable into segregated and functionally independent modules, Pessoa asks us to consider the very minimal operations a brain requires to enable an animal to visually perceive threats to fend off and rewards to seek. Such a 'minimal brain' would likely include the optic tectum—an ancient structure that, in mammals, evolved to become the superior colliculus. Located in the 'roof' of the midbrain ('tectum' is roof in Latin), this structure was traditionally associated with visual integration in non-mammalian brains. But, as we learn in chapter 3, further research has revealed that in mammalian—and even in some non-mammalian—brains, this intriguing brain structure is highly sensitive not only to other kinds of sensory information but to internal and somatic information as well. This 'context sensitivity', Pessoa argues, is 'a fundamental principle of brain function: The brain does not simply react to sensory stimuli; instead, incoming data are incorporated into ongoing processing that encompasses the states of the brain and the body, explaining why the exact same stimulus exerts very different effects depending on the situation: in one setting a stimulus may lead to inquisitive approach, in another to moving away' (pp. 37–38).

In addition to context sensitivity, brain areas also exhibit multifunctionality. According to Pessoa, the view of the brain as a modular system is tightly associated with the assumption that individual brain areas have distinct functions. However, as he argues in chapter 4, the evidence overwhelmingly suggests that the mapping from brain areas to psychological functions is not one-to-one but rather many-to-many. Indeed, as he reviews in this chapter, Pessoa himself has conducted important work in neuroimaging showing that different brain regions are associated with multiple psychological functions, to different degrees, and that the extent of their involvement also depends on the particular task employed in each study (Anderson et al. [2013]). Thus, instead of talking about the function of a brain region, Pessoa introduces the notion of 'functional profile' to indicate that each brain region is more or less involved in different functions, and that such profiles are likely dynamic and task dependent.

The context sensitivity and multifunctionality of brain areas is thoroughly explored in chapters 5 and 6, where Pessoa explores the neural intricacies underlying emotion and motivation. In these chapters, we learn not only how entangled the subcortical (chapter 5) and cortical (chapter 6) pathways supporting these processes are, but also that no single area is the 'neural protagonist' of a cognitive operation—that is, that the amygdala is not the 'organ of emotion', say, nor the hippocampus 'the seat of memory'. Moreover, in chapter 7, Pessoa explores how this interconnectedness pervades even higher-order cognitive processes, such as cognitive control, that have been traditionally associated with the prefrontal cortex. As it happens, even in prototypical tasks that are supposed to involve only cognitive control, specific areas turn out to be supported by multiple bidirectional connections that go beyond the confines of a single brain lobule. Taken together, then, these three chapters seek to demonstrate how difficult, if not impossible, it is to fit familiar cognitive processes such as emotion, motivation, and cognitive control within the hierarchical, reductionistic formula we have been sold for years in the cognitive neurosciences.

The suggestion is then to abandon the hierarchical, reductionistic, and modular view of the brain and to think of it instead as a complex system. Thus, in chapter 8, Pessoa gives us a brief introduction of the notion of complex systems, highlighting the need to understand them as dynamic and non-linear, which—according to him—goes contrary to the standard, mechanistic way of seeing the brain as modular and decomposable. At the end of this chapter, we are left with six essential implications that follow from thinking of the brain as a complex system:

(1) The brain is a system of interacting parts, linked by anatomical connections and not a set of independent functionally localizable units.

(2) The notion of 'levels of analysis' should be understood in epistemic and pragmatic terms, rather than as a metaphysical thesis of the functional organization of the human brain.

(3) The brain isn't a static organ and its processes aren't simple linear cause–effect interactions.

(4) The brain is not a centralized hierarchy, but rather a distributed heterarchy.

(5) Emergent properties are the norm in complex systems such as the brain.

(6) The brain is complex, in the sense of being erratic and unstable, to the point that many of its operations are likely unpredictable.

Why would our cognitive, affective, and perceptual capacities depend on such a complex organ? To answer that question, in chapter 9 Pessoa offers a beautiful excursus into the last 500 million years of evolution of the brain. We learn why different structures evolved long reaching connections from subcortically and cortically distant areas, why certain loops among these connections emerged, and—critically—why it is that we should not think of brain evolution as simply a matter of building new structures upon old ones. Regions we share with common ancestors continued to evolve independently and to create their own reciprocal connections, making it hard to ascertain the precise sense in which a brain structure is 'preserved'. Indeed, this chapter is a must-read for anyone interested in the notion of homology in the brain, and in the extent to which it is safe to extrapolate to humans neuroscientific results that come from research in non-human animals.

Up until this point, Pessoa has argued against the prevailing view in cognitive neuroscience that takes the brain as a hierarchical, decomposable, and modular system, and he has argued instead for a view of the brain as a heterarchical, dynamical, and complex system. Does this mean that cognitive neuroscience is doomed? Not necessarily. Chapter 10, which in a sense constitutes the heart of Pessoa's positive proposal, offers a way forward. In essence, his suggestion is that we should think of cognitive, affective, and perceptual processes as emerging from the dynamic activity of the brain's large-scale networks forming 'functionally integrated systems': 'The brain's anatomical and functional architectures are highly nonmodular; signal distribution and integration are the norm, allowing the confluence of information related to perception, cognition, emotion, motivation, and action; and the functional architecture is composed of overlapping networks that are highly dynamic and context-sensitive' (p. 168).

According to Pessoa, functionally integrated systems are based on five organizational principles: (1) massive combinatorial anatomical connectivity (in the sense that there are multiple connections between nodes and several ways to get from one node to another); (2) highly distributed functional connectivity, with the proviso that the connections between regions may not only be non-binary but may not even be captured by a single weight value; (3) networks as functional units, as opposed to brain areas (and, one may suspect, individual neurons); (4) interactions via cortical–subcortical loops; and (5) connectivity with the body, for neuronal connections extend beyond the confines of the cranium—the brain's functionally integrated systems are embedded. It is thanks to these large-scale networks, these functionally integrated systems, that we have the cognitive, affective, and perceptual abilities we enjoy.

To explain how thinking of the brain in terms of functionally integrated systems can help us to understand the kinds of complex behaviours it supports, Pessoa offers, in chapter 11, a concrete example: extinction learning. How can the brain manage to unlearn a previous association between an unconditioned stimulus (for example, a light) and a conditioned stimulus (for example, a shock)? What we learn in this chapter is that the process is complex, not only because it involves rather intertwined circuitry, but also because the underlying mechanisms that supposedly code for variables such as value or reward are context-sensitive and can only be understood in relation to the operations of other connected areas. The book ends, in chapter 12, with provocative and brief discussions of some of the consequences that would follow from thinking of the brain as a complex system composed of dynamic and functionally integrated networks. Such consequences include rethinking the nature of explanation in cognitive neuroscience, expanding our repertoire of what constitutes causation in the brain, and even reconsidering our ontological commitments so that we take the basic building blocks of reality to be processes rather than things.

I expect that some philosophers of mind and neuroscience may find this book, at times, frustrating. After all, we philosophers are experts in identifying ambiguities, equivocations, and conceptual imprecisions. And Pessoa has left open plenty of opportunities for us to show our skills. For instance, I can see how one could argue that modularity need not stand in opposition to topological approaches in brain science (Stanley and De Brigard [2016]; Stanley et al. [2019]). Likewise, as Boone ([2024]) recently showed, it is possible for reciprocally interconnected structures to nonetheless be amenable to analysis via functional decomposition. It is even possible that thinking of the brain as a complex dynamical system won't be at odds with mechanistic explanations, as suggested by Barack ([2021]). Moreover, there is plenty of room for rejecting what Pessoa calls 'the axiom of biology' without having to embrace multifunctionality, modularity, or one-to-one mappings from brain structures to cognitive operations (Boone et al. [2023]). All of these are conceptual moves that are left open for the curious reader.

I am not interested, however, in highlighting any of these issues in this review. Instead, I want to end by praising Pessoa's book for showing how the more we discover about the brain, the more challenging the simple layer cake view of neuro-reductionism becomes (Oppenheim and Putnam [1958]; Wimsatt [1976]). I doubt that the right approach to understanding the brain would require us to fully jettison every attempt to understand any brain area as a single unit, but I also doubt that anatomical or functionally parcellated structures neatly map onto the kinds of cognitive functions that form the table of contents of our contemporary textbooks in neuroscience. Likewise, I doubt that network models produced by our current neuroimaging techniques would unproblematically reveal the correct neural correlates of a future depurated cognitive ontology, but I also doubt that we can make any progress in understanding how the brain relates to the mind if we keep on studying brain structures in isolation. Either way, one thing is clear:

we won't be able to make much progress without learning more about how the brain is profoundly interconnected at different scales, and without occasionally revising some of our most cherished neuroscientific dogmas. And these are sufficient reasons for every philosopher interested in the mind and in neuroscience to read this stimulating book.

Felipe De Brigard
Duke University
felipe.debrigard@duke.edu

References

Anderson, M. L., Kinnison, J. and Pessoa, L. [2013]: 'Describing Functional Diversity of Brain Regions and Brain Networks', *Neuroimage*, **73**, pp. 50–58.

Barack, D. L. [2021]: 'Mental Kinematics: Dynamics and Mechanics of Neurocognitive Systems', *Synthese*, **199**, pp. 1091–123.

Boone, T. [2024]: 'Robustness and Modularity', *British Journal for the Philosophy of Science*, **75**, pp. 417–42.

Boone, T., Van Rooy, N. and De Brigard, F. [2023]: 'Not Every Thing Must Go', *Journal of Cognitive Neuroscience*, **35**, pp. 376–79.

Kaku, M. [2014]: *The Future of the Mind: The Scientific Quest to Understand, Enhance, and Empower the Mind*, New York: Double Day.

Oppenheim, P. and Putnam, H. [1958]: 'Unity of Science as a Working Hypothesis', *Minnesota Studies in the Philosophy of Science*, **2**, pp. 3–36.

Stanley, M. and De Brigard, F. [2016]: 'Modularity in Network Neuroscience and Neural Reuse', *Behavioral and Brain Sciences*, **39**, available at .

Stanley, M. L., Gessell, B. S. and De Brigard, F. [2019]: 'Network Modularity as a Foundation for Neural Reuse', *Philosophy of Science*, **86**, pp. 23–46.

Wimsatt, W. C. [1976]: 'Reductionism, Levels of Organization, and the Mind–Body Problem', in G. Globus, G. Maxwell and I. Savodnik (*eds*), *Consciousness and the Brain: A Scientific and Philosophical Inquiry*, New York: Plenum, pp. 205–67.

Recent Reviews

Matteo Mameli, Why Human Nature Matters // Marco J Nathan, The Quest for Human Nature // Reviewed by Tim Lewens

Hans-Jörg Rheinberger, Split and Splice // Reviewed by Uljana Feest

Marshall Abrams, Evolution and the Machinery of Chance // Reviewed by Charles H Pence