The (Un)Scientific Status of Integrated Information Theory

Tony Cheng
Waseda Institute for Advanced Study

Recently, a group of 124 scientists and philosophers wrote a public statement to express serious concerns about the integrated information theory of consciousness (IIT for short), especially about how it has been falsely promoted as a scientifically established theory. Some points made in the statement, especially the use of the term "pseudoscience," have caused what has been described as an "uproar" on social media. This is a prime example of philosophy of science going public, but perhaps the negative impact is stronger than the positive. In this talk, I will selectively summarise the development of this debate concerning IIT's scientific status, and gesture ways of moving forward in this debate of philosophy of cognitive neuroscience.

Initially proposed by the neuroscientist Giulio Tononi (2004), the core of IIT has it that consciousness is identical to a certain kind of information, the realisation of which requires integration that can be measured mathematically according to the now famous *phi* metric. More specifically, IIT takes consciousness as primary in that it cannot be analysed. However, the theory advances five axioms that are supposed to capture the nature of consciousness. They are axioms in that these dimensions of conscious experiences are self-evident. They include the idea that consciousness is real and undeniable in an intrinsic way; the idea that consciousness has composition in that each conscious experience has a specific structure; the idea that information can distinguish one experience from other experiences; the idea that consciousness is irreducible to separate elements and therefore unified; and finally the idea that conscious experiences specify certain things and thereby exclude other things. In addition to these axioms, IIT also has a number of postulates and empirical predictions. It has been updated in the past two decades, with the newest version known as IIT 4.0 (Albantakis et al., 2023).

Although the above features seem to make IIT as a philosophical, as opposed to scientific theory (i.e., having axioms, postulates, etc.), IIT does claim itself as scientific, and the proponents often publish papers in scientific journals (e.g., Barbosa et al., 2020; Marshall et al., 2023; Cogitate Consortium, et al., 2025). The key of IIT's scientific status lies in the fact that it seems to make clear empirical predictions. In particular, it makes predictions about the neural correlates of consciousness (NCC). The crucial divide here is between the anterior and the posterior theory, in that they disagree about NCC's location in the brain. IIT requires that the so-called "posterior cortical hot zone" – the specific parts of neocortex – sustains the relevant kind of consciousness (Koch, Massimini, Boly, and Tononi, 2016). Why isn't this enough to guarantee the scientific status of IIT?

The talk will focus on the crucial fact that IIT is in effect a version of panpsychism in metaphysics, and therefore unscientific. John Searle (2013) and David Chalmers (2016) have pointed this out long before, but whether this makes IIT pseudoscientific remains unclear. I will here argue that IIT is indeed unscientific even if it seeks to make concrete empirical predictions about NCC.