

Inputs, Outputs, and Meta-Models

Aïda Elamrani

Institut Jean Nicod, ENS

The talk proposes a meta-theoretical way to compare scientific theories of consciousness by treating them as functions: systems with inputs, outputs, and rules that map one to the other. Its starting point is the difficulty of defining consciousness. Researchers from neuroscience, computer science, psychology, philosophy, and mathematics often mean different things by the term: agency, selfhood, qualia, phenomenal properties, intentionality, or access to mental contents. Because consciousness studies is becoming a scientific field, I argue that it needs a shared language capable of bridging these disciplinary backgrounds.

The historical background is the old mind/matter problem. Earlier ideas about souls and post-mortem existence have been replaced by increasingly formal and mathematical ways of describing mind: statistics in psychology, behaviorism, computational models, EEG signatures, and even quantum approaches. But mathematical sophistication alone does not solve the conceptual problem. Paradoxically, a scientific explanation of consciousness is expected to explain it through its physical substrate.

Mathematics matters here because they provide a formal language: symbols, syntax, and rules that allow precise communication and prediction. But if we stay too close to specific mathematical formalisms, we may lose sight of the larger comparative framework. Nearly any mathematical model can be viewed as a function: an input is processed through rules or an algorithm to produce an output. This simple structure can help clarify what a theory of consciousness is actually claiming.

The talk distinguishes this proposal from classical functionalism, especially the version associated with Putnam and early artificial intelligence. That classic view treated mental states as functional/computational states defined by inputs, outputs, and transitions. It ran into well-known problems: it seemed unable to account for phenomenology, qualia, intentionality, and reference to objects in the world. I am not reviving functionalism as a theory of mind, but using “function” as a neutral meta-model for comparing theories.

The proposed “puzzle” is to ask where consciousness sits in the function. It could be treated as the input, as in theories that begin with measurements of consciousness and produce a value, such as IIT’s phi, or even a yes/no classification. It could be treated as the output, where physical data are mapped to a conscious state, resembling the brain/software analogy. Or consciousness could be treated as the function itself: an interface or process that maps physical data into a “mind space,” broader than just qualia because minds include more than phenomenal properties.

The payoff of this framework is diagnostic. For any theory, it forces us to specify the input space, output space, mapping rules, what is computable, what is excluded, and how physical and mental descriptions are related. Under this view, the question is not whether the brain is a computer, but what kind of formal relation between matter and mind a theory of consciousness is proposing.