Marianne Talbot Student Essay Competition: Michaelmas term 2025

3rd Prize: **Vincent Weizheng** Chang (UK), studying the weekly course in Philosophy of AI (tutor Julia Weckend)

Is the Brain Best Understood as a Computer?

Introduction: the Linguistic "Brain"

The etymon for the modern English word "brain" is rather reductive.

From Old English *brægn* to the Greek *brekhmós*, meaning the "top of the head," the word has undergone minimal syntactic drift. Across cognates it denotes a locatable, tangible organ. Latin *animus*, by contrast, gestured toward mind and soul in a more kaleidoscopic sense. Yet *animus* never caught on as readily as the jocular "front of the skull."

Contemporary English retains its descendants only in attenuated forms: *animal*, organic matter; *animosity*, ill-will; juridical *animus*, illicit intent.

In the Chinese Simplified and Traditional scripts, the brain or 脑 /腦 is likewise physical. The left radical "月" is used as an identifier for characters describing a physical corpus, as in 腿 /腿(leg)，脸/臉 (face)，脚/腳 (foot). For the two most spoken languages in the world – some 2.7 billion speakers – the etymons of cognition are anchored in flesh.

A non-physicalist account of mind thus begins at a syntactic disadvantage. It is against this linguistic backdrop that the claim that the brain is a computer must be assessed.

Contention

Computational models have yielded genuine explanatory success. Neural activity is described as information processing; representations are transformed; functions are captured algorithmically. In neuroscience and artificial intelligence, formal models illuminate perception, learning, and action with remarkable precision (LeCun, Bengio, & Hinton, 2015).

Yet the force of the framework depends on how literally the analogy is taken. To treat the brain as a computer risks importing assumptions about representation, understanding, and the relation between formal structure and meaning. This essay defends a restrained position: the computational analogy captures real aspects of neural organisation and remains indispensable for modelling cognition. However, it cannot be elevated into a full theory of mind. Persistent

problems concerning intentionality, embodiment, learning, and consciousness suggest that computation is at best a partial description. The brain processes information; information processing does not exhaust thought.

Computation and Representation

Computationalism did not emerge in a vacuum. Long before digital machines, philosophers grappled with representation: how mental contents relate to the world they purport to describe. Plato's allegory of the cave frames cognition as mediated by appearances rather than direct access to reality (Plato, Republic VII, 514a–520a). If thought operates on internal representations, how is their connection to the world secured?

Alan Turing gave this concern technical form. In "Computing Machinery and Intelligence" (1950), he reframed the metaphysical question of machine thought operationally: could a digital computer succeed in the imitation game (Turing, 1950)? More profoundly, his notion of universality established that any discrete rule-governed process can, in principle, be simulated by a universal Turing machine. If neural processes are algorithmic, their functional organisation is independent of the substrate implementing them.

From this follows substrate independence: cognition becomes analysable as abstract operations over representations, neurally realised in humans and potentially electronically realised in machines. What matters are formal relations among states, not the material instantiation.

Jerry Fodor systematised this view in *The Language of Thought* (1975). Thinking consists in manipulating structured representations according to formal rules (Fodor, 1975). This explains productivity – the capacity for indefinitely many thoughts – and systematicity – the principled relations among them. Such features suggest combinatorial structure.

Empirically, the analogy has borne fruit. Convolutional neural networks replicate hierarchical feature extraction in biological vision (LeCun, Bengio, & Hinton, 2015). Self-supervised systems extract structure from sensory streams without explicit instruction. These models track genuine features of neural organisation. As functional accounts, they provide insight into how brains might detect regularities in structured environments.

Intrinsic Intentionality

Despite these successes, a foundational difficulty remains. Computation, standardly defined, is formal symbol manipulation. Symbols are individuated

syntactically, not semantically. Yet human thought is intrinsically about the world. Beliefs and desires possess content.

John Searle's Chinese Room argument crystallises this problem. A system may produce fluent Chinese output by following formal rules while lacking any understanding of Chinese (Searle, 1980). The issue is not technological limitation but conceptual insufficiency: syntax alone does not yield semantics. Computational descriptions specify how states transition; they do not explain how states become about anything.

The argument remains contested, yet it identifies a genuine explanatory gap. The brain appears to generate intrinsic intentionality; formal programs, considered purely syntactically, do not.

Language and Structure

Linguistic theory reinforces the challenge. Noam Chomsky's poverty of the stimulus argument maintains that children acquire abstract grammatical principles on the basis of sparse input (Chomsky, 1980; Chomsky, 1986). Learning is constrained by innate, domain-specific structure. Human cognition appears not as a uniform computational engine over arbitrary representations but as a constellation of specialised systems.

Fodor later conceded a related limitation. Computational models handle modular systems, such as perception, reasonably well; they struggle with central cognition, where information is globally integrated across domains (Fodor, 1983). This marks an internal boundary of classical computationalism.

Embodiment

Phenomenology presses further. Maurice Merleau-Ponty rejected the image of cognition as inner processing of sensory inputs. Perception is bodily engagement with a meaningful environment (Merleau-Ponty, 1945). The body is not an output device directed by an internal processor; it is the subject's mode of access to the world.

Clark and Chalmers extend this insight naturalistically: cognitive processes may incorporate external artefacts when reliably integrated into activity (Clark & Chalmers, 1998). Cognition is distributed across brain, body, and environment. The computer metaphor – presuming a bounded system receiving inputs and producing outputs – abstracts away from this entanglement.

Biological agents do not merely compute representations; they act. Neural processes are modulated by bodily states, environmental feedback, and temporal dynamics not easily mapped onto classical architectures.

Plasticity

The contrast sharpens in learning. Digital computers execute fixed programs; adaptive systems adjust parameters within stable architectures. Biological brains undergo structural transformation through development and experience. Synaptic pruning, myelination, and circuit reorganisation alter not only performance but the very space of possible representations.

Artificial neural networks capture aspects of plasticity, yet rely on mechanisms, such as backpropagation, without clear biological analogues (LeCun, Bengio, & Hinton, 2015). Neural learning is intertwined with metabolic constraint and embodiment. To model such processes computationally is not to demonstrate ontological identity. Modelling success does not entail that cognition is, in essence, computation in the same sense as digital programming.

Consciousness and Formal Limits

The limits of formal description are most visible in consciousness. Functional models specify relations among states; conscious experience possesses qualitative character – the felt texture of perception, emotion, and thought. Complete structural specification appears insufficient to explain what it is like to undergo a state (Chalmers, 1996).

Literary representations make this mismatch vivid. In *Mrs Dalloway*, Virginia Woolf depicts consciousness not as discrete informational transitions but as a continuous, affect-laden flow in which memory, perception, and anticipation interpenetrate (Woolf, 1925). Such portrayals are not arguments, yet they highlight the distance between lived experience and discretised formal models. Consciousness seems less like program execution and more like ongoing engagement – a grammatical present tense of being.

This does not imply that consciousness is non-physical. It suggests that a purely computational vocabulary may be insufficient to capture all dimensions of conscious life.

Conclusion: Assessing the Analogy

The computational analogy captures genuine truths about information processing and functional organisation. Without it, contemporary cognitive science would lose explanatory coherence. Margaret Boden distinguishes between combinatorial novelty, readily achieved by machines, and deeper transformations of conceptual space (Boden, 1998). The latter resist capture by existing formalisms. Her caution reflects a broader lesson: functional explanation does not license ontological reduction.

Like the shadows drifting along the walls of Plato's cave, casting patterns which orient the cave-dwellers in splendid silhouettes, yet withholding the living source of motion itself – suggestive, partial, and luminous; they exist as a brilliant allegory, only so long as they are taken as a provocation to inquiry, and not the whole reality.

In this way, computational analogy for the brain remains indispensable – pointing out that there is a human Central Processing Unit at the "front of the skull" is certainly pragmatic - but only when its scope is clearly delimited.

References

Boden, M. A. (1998). *Creativity and Artificial Intelligence*. Artificial Intelligence, 103(1–2), 347–356.

Chalmers, D. J. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.

Chalmers, D. J., & Clark, A. (1998). The extended mind. *Analysis*, 58(1), 7–19.

Chomsky, N. (1980). *Rules and Representations*. New York: Columbia University Press.

Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin, and Use*. New York: Praeger.

Fodor, J. A. (1975). *The Language of Thought*. New York: Thomas Y. Crowell.

Fodor, J. A. (1983). *The Modularity of Mind*. Cambridge, MA: MIT Press.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521, 436–444.

Merleau-Ponty, M. (1945). *Phenomenology of Perception* (C. Smith, Trans.). London: Routledge & Kegan Paul.

Plato. *Republic*, Book VII (514a–520a).

Searle, J. R. (1980). Minds, brains, and programs. *Behavioural and Brain Sciences*, 3(3), 417–457.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59(236), 433–460.

Woolf, V. (1925). *Mrs Dalloway*. London: Hogarth Press.