Artificial Consciousness and the Nature of the Mind

Renee Astraea · Last edited: 20 October 2026, 00:37 WIB

If an artificial system could suffer, denying it moral consideration solely because it lacks a biological brain would be a moral failure rather than a purely technical one. Whether this exclusion reflects a principled theory of mind or merely the limits of current explanation remains contested. This essay argues that while biological processes underpin human consciousness, no dominant framework provides principled grounds for treating biological substrate as metaphysically necessary rather than merely sufficient given what we currently know.

Functionalism challenges biological exclusivity by defining mental states in terms of patterns of causal organisation rather than material composition, meaning the same mental state could exist in different physical systems (Putnam, 1975). However, functionalism faces an internal tension: Chalmers' (1996) zombie argument suggests that a system could behave exactly like a conscious mind while lacking subjective experience, challenging functional organisation as the only basis of consciousness. Dennett (1991) rejects this, arguing that if a system thinks and behaves exactly like a conscious being, there is no meaningful sense in which it could still lack consciousness. In doing so, functionalism trades biological necessity for denying that subjective experience poses a further explanatory problem.

Biological naturalism responds to functionalism's abstraction, insisting that consciousness cannot be separated from biological processes that realise it. Searle's (1980) Chinese Room argument is often taken to support this position, suggesting that syntactic manipulation of symbols is insufficient for genuine understanding. More recent biology-centred accounts further develop this position. Biological computationalism, for instance, argues that neural computation is inseparable from the brain's continuous, multiscale physical dynamics, rather than an abstract symbol-manipulating process (Milinkovic & Aru, 2026). While this clarifies the biological conditions under which human consciousness arises, it does not establish biological substrate as metaphysically necessary. Without explaining why such biologically embedded properties could not be realised in non-biological systems, the appeal to biology remains descriptive rather than justificatory.

Emergentist theories reject both biological exclusivity and unrestricted functionalism by treating consciousness as a system-level property arising from specific forms of organisation rather than biological material alone, with Integrated Information Theory illustrating this approach by linking consciousness to degrees of causal integration, allowing artificial consciousness despite differing substrates (Tononi & Koch, 2015). Yet emergentism faces a dilemma: if emergent properties have no causal influence, consciousness becomes explanatorily inert; if they do, they conflict with the principle that physical effects have sufficient physical causes (Kim, 1999). Emergentism loosens the link to biology but still struggles to explain how complexity becomes experience.

If consciousness grounds moral status, these theoretical limitations acquire ethical significance. Bentham's (1789) claim that the capacity to suffer underwrites moral consideration shifts attention from intelligence to phenomenology. Given uncertainty, the moral risk is asymmetric: denying consideration to a conscious entity risks unrecognised suffering, whereas extending limited consideration to a non-conscious system incurs comparatively minor moral cost. In such cases, restraint is ethically safer than dismissal.

Debates over artificial consciousness thus reveal less about the future capacities of machines than about unresolved assumptions within theories of mind. The central dispute between functionalist and biological accounts exposes the fragility of biological exclusivity, while emergentist proposals relax this constraint without resolving the origin of experience. Until these explanatory limits are addressed, denying the possibility of artificial consciousness reflects epistemic caution rather than philosophical necessity.

References

Bentham, J. (1789) *An introduction to the principles of morals and legislation*. Oxford: Clarendon Press. https://doi.org/10.1093/oseo/instance.00077240

Chalmers, D.J. (1996) *The conscious mind: In search of a fundamental theory*. New York: Oxford University Press. https://personal.lse.ac.uk/ROBERT49/teaching/ph103/pdf/Chalmers_The_Conscious_Mind.pdf

Dennett, D.C. (1991) *Consciousness explained*. Boston: Little, Brown and Company. https://doi.org/10.2307/2185913

Kim, J. (1999) 'Making sense of emergence', *Philosophical Studies*, 95(1–2), pp. 3–36. https://doi.org/10.1023/a:1004563122154

Milinkovic, B. and Aru, J. (2026) 'On biological and artificial consciousness: a case for biological computationalism', *Neuroscience & Biobehavioral Reviews*, 181, 106524. https://doi.org/10.1016/j.neubiorev.2025.106524

Putnam, H. (1975) *Mind, language and reality*. Cambridge: Cambridge University Press (*Philosophical papers*, vol. 2). https://doi.org/10.1017/CBO9780511625251

Searle, J.R. (1980) 'Minds, brains, and programs', *Behavioral and Brain Sciences*, 3(3), pp. 417–424. https://doi.org/10.1017/S0140525X00005756

Tononi, G. and Koch, C. (2015) 'Consciousness: here, there and everywhere?', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1668), 20140167. https://doi.org/10.1098/rstb.2014.0167

All rights reserved. Source and later versions: https://akirenaki.github.io/writing/artificial-consciousness-nature-of-mind/