Consciousness science as a marketplace of rationalizations

Matan Mazor, All Souls College and Department of Experimental Psychology, University of Oxford, UK; matan.mazor@psy.ox.ac.uk; matanmazor.github.io

Commentary on Fleming, S. M., & Michel, M. (2024). Sensory horizons and the functions of conscious vision. *Behavioral and Brain Sciences*, 1-53.

In press, *Behavioral and Brain Sciences*

Abstract: In consciousness science, theoretical predictions are often untestable, such as claims about phenomenal consciousness in other beings. This evidential underdetermination, in combination with the perceived moral significance of consciousness, puts consciousness science at risk of becoming a marketplace of rationalizations: a field that produces theories that reaffirm social practices and conventions.

What makes a good scientific theory of consciousness? As a starting point, it needs to agree with the available evidence from behaviour, physiology, and introspection, and be consistent, both internally and with other accepted theories (for example, with evolution by natural selection). Other desired characteristics are breadth of scope, simplicity, and the ability to generate new research directions (Kuhn, 1977). While it is too early to judge whether the proposal by Fleming and Michel (this volume) agrees with all available evidence, it is certainly ambitiously broad, simple, and generative. Motivated by evidence for the slowness of visual conscious experience under some conditions, Fleming and Michel identify the emergence of consciousness in evolution with the development of long-distance vision and model-based planning, which, following MacIver, they trace back to the water-to-land transition, around 400 million years ago.

In addition to Kuhn's five abovementioned epistemic values of accuracy, consistency, breadth, simplicity and generativity, their theory scores high according to an additional metric that is easy to miss: it aligns with many readers' pre-scientific intuitions about *morality* — specifically about the moral status of fish. Fleming and Michel argue that most fish are not phenomenally conscious (there is no "something it is like" to be them, at least not in terms of their visual experience), because the murky and dark water of the deep sea never allowed them to develop a capacity to plan into the future. A theory that reveals that fish are unconscious zombies will not offend many readers (how many of us are emotionally attached to a pet fish?), and will be received as good news by those readers who have a taste for seafood.

Viewed this way, theory selection in consciousness science can be described as operating within a *marketplace of rationalizations* (Williams, 2023). Such marketplaces emerge because people are often motivated to hold beliefs for reasons outside their instrumental value. We like to believe that our actions are morally justified, that we are liked and respected by our peers, and that we will live a long and healthy life. And yet, it is psychologically impossible to willfully decide to believe something because it feels nice: knowing that you had a non-epistemic motivation for a belief would undermine your ability to truly hold it (Williams, 1973). Humans are therefore put in a peculiar position in which they try to deceive themselves into thinking that beliefs they hold for hidden, non-epistemic reasons, are held for purely epistemic reasons. To pull this off, argues Williams, people are in constant search of rationalizations: pieces of seemingly objective information that support their desired beliefs. This makes such rationalizations an economic good that can be traded for money, social status, and, in an academic context, publications, grant money, and citations.

Two facts about consciousness science make it particularly conducive to such a marketplace. First, consciousness is central to lay conceptions of morality (Gray et al., 2012; Mazor et al., 2021; Hirschhorn et al., 2025). To a good first approximation, people care about others to the extent that they think there is “something it is like” to be them. This, together with a desire not to be perceived (by others and by oneself) as someone who harms others, produces a motivation to only attribute consciousness to beings that are currently treated with care and respect (Mazor et al., 2023). Such motivated reasoning can be experimentally demonstrated in the lab: in one study, participants attributed lower levels of mental capacities to sheep and cows if they believed they would later eat meat, compared with fruit (Bastian et al., 2012).

This market demand for rationalizations is a first important factor, but in itself it is not sufficient to form a marketplace. After all, people are also motivated to believe they will live a long and healthy life, but scientists’ ability to produce theories that carry good news about life expectancy is importantly limited by hard facts from scientific observations. In consciousness science, however, theory is radically underconstrained by evidence. No evidence can prove that a fish has, or does not have, a subjective point of view. We are left with a field that produces theories with potentially major implications for ethics and policy but without any ability to measure the phenomenon itself (in this case, the subjective experience of being a fish). Under these conditions, whether a theory of consciousness can produce better rationalizations (for example, rationalizations for the practice of killing and eating fish) will be a primary determinant of its success.

The human need for rationalizations has shaped consciousness science since its inception. Descartes saw his view that animals were unconscious automatons “not so much cruel to beasts but respectful to human beings… whom it absolves from any suspicion of crime whenever they kill or eat animals” (Descartes, 1999, cited in Kaldas, 2015). More recently, a new wave of biological naturalism, arguing that consciousness is inherently a property of living beings (Aru et al., 2023; Findlay et al., 2024; Seth, 2025) coincided with concerns about the potentially devastating moral implications of conscious AI (Long et al., 2024). As with Fleming and Michel’s fish, or Descartes’ beasts, there may be good epistemic reasons to favour a theory in which AI cannot be conscious. Yet a full understanding of consciousness science requires taking seriously the force of rationalizations in theory formation and selection.

A marketplace of rationalizations is a system-level description rather than an account of the conscious intentions of individual actors. Scientists typically assume that theories are valued primarily based on epistemic merits. This is in contrast to more canonical examples of marketplaces of rationalizations, such as politicised media, where content producers may willfully generate material that appeals to consumers' non-epistemic motivations. Yet given that consumers (grant agencies, journals, popular media, and fellow scientists) *do* weigh theories by their non-epistemic merits too, and since scientists themselves are often implicitly motivated to rationalize beliefs they hold for non-epistemic reasons, a marketplace of rationalizations remains a useful model for consciousness science. As with the development of planning in early terrestrial animals, opening our eyes to the structure of this market may allow scientists to better navigate it.

Aru, J., Larkum, M. E., & Shine, J. M. (2023). The feasibility of artificial consciousness through the lens of neuroscience. *Trends in neurosciences*, *46*(12), 1008-1017.

Bastian, B., Loughnan, S., Haslam, N., & Radke, H. R. (2012). Don't mind meat? The denial of mind to animals used for human consumption. *Personality and Social Psychology Bulletin*, *38*(2), 247-256.

Descartes, R. (1999). *Meditations and other metaphysical writings*. Penguin.

Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological inquiry*, *23*(2), 101-124.

Findlay, G., Marshall, W., Albantakis, L., David, I., Mayner, W. G., Koch, C., & Tononi, G. (2024). Dissociating artificial intelligence from artificial consciousness. *arXiv preprint arXiv:2412.04571*.

Hirschhorn, R., Negro, N., & Mudrik, L. (2025). Does consciousness matter morally? A survey of folk and expert intuitions. A poster presented at the annual meeting of the *Association for the Scientific Study of Consciousness*

Kaldas, S. (2015). Descartes versus Cudworth On The Moral Worth of Animals. *Philosophy Now. https://philosophynow.org/issues/108/Descartes_versus_Cudworth_On_The_Moral_Worth_of_Animals*

Kuhn, T. S. (1977). Objectivity, value judgment and theory choice. In.: The essential tension. Urbana: University of Illinois Press.

Long, R., Sebo, J., Butlin, P., Finlinson, K., Fish, K., Harding, J., ... & Chalmers, D. (2024). Taking AI welfare seriously. *arXiv preprint arXiv:2411.00986*.

Mazor, M., Risoli, A., Eberhardt, A., & Fleming, S. M. (2021). Dimensions of moral status. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 43, No. 43).

Mazor, M., Brown, S., Ciaunica, A., Demertzi, A., Fahrenfort, J., Faivre, N., ... & Lubianiker, N. (2023). The scientific study of consciousness cannot and should not be morally neutral. *Perspectives on Psychological Science*, *18*(3), 535-543.

Seth, A. K. (2024). Conscious artificial intelligence and biological naturalism. *Behavioral and Brain Sciences*, 1-42.

Williams, B. (1973). Deciding to believe.