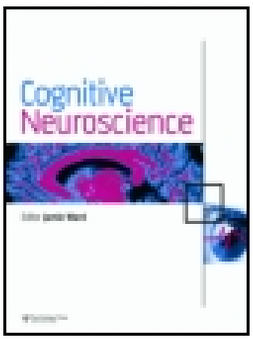

Routledge
Taylor & Francis Group

Cognitive Neuroscience

Current Debates, Research & Reports

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/pcns20

Theories of consciousness are solutions in need of problems

Stephen M. Fleming

To cite this article: Stephen M. Fleming (2020): Theories of consciousness are solutions in need of problems, Cognitive Neuroscience, DOI: 10.1080/17588928.2020.1841744

To link to this article: https://doi.org/10.1080/17588928.2020.1841744

COGNITIVE NEUROSCIENCE
https://doi.org/10.1080/17588928.2020.1841744

Routledge
Taylor & Francis Group

Theories of consciousness are solutions in need of problems

Stephen M. Fleming

Wellcome Centre for Human Neuroimaging and Department of Experimental Psychology, University College London, London, UK

ABSTRACT

Doerig et al. point out a number of shortcomings with existing theories of consciousness and argue they should be systematically constrained by empirical data. In this commentary I suggest a further constraint – the potential functions of (the contents of) consciousness. One such candidate function in humans is the social sharing of reportable mental states. The social function of consciousnessprovides a general framework within which to understand the evolution and neurobiology of conscious awareness.

ARTICLE HISTORY
Received 1 September 2020
Revised xx xxx xxxx
Published online xx xxx xxxx

KEYWORDS
consciousness; computation; social function

A common approach in consciousness science is to start with subjective experience and then attempt to bridge an 'explanatory gap' from experience to theoretical models or neural substrates. It is becoming increasingly clear that this approach to the problem of consciousness is in danger of losing its way, as nicely highlighted by Doerig's target article. They point out a number of shortcomings with existing theories of consciousness, and in particular emphasize the need for 'hard criteria' on what constitutes a useful theory. In this commentary I suggest a further constraint on these criteria – the functions of (the contents of) consciousness.

A useful framework for thinking about the function of computational systems is David Marr and Tomaso Poggio's influential taxonomy of three distinct levels of description (Marr & Poggio, 1976). A computational level characterizes the problem to be solved in terms of available inputs and desired outputs. An algorithmic level then specifies how this computation is performed – as a system of mathematical equations, lines of code or diagrams of interacting variables. Finally, an implementational level specifies how our chosen algorithm is implemented in hardware or wetware – for instance, how a computer program is implemented in the machine code of your laptop, or how dark adaptation is achieved by cells in the retina.

The definition of Marr and Poggio's uppermost level as 'computational' can be confusing to contemporary readers as the term computation is now often used to refer to different algorithms (such as reinforcement learning) rather than the problem to be solved (seeking food rewards). For the purposes of this commentary, let's define this level as one identifying a putative (evolved) function, or the functional level for short. The paradigm cases of consciousness highlighted by Doerig et al. (such as visual masking) represent a significant step toward identifying such functions, as they aim to cleanly separate conscious and unconscious alternatives – such as being able to report we are experiencing contents *X*, versus not being able to do so. However, for the most part they leave the functional level unexplored – why might it be useful for a system to report that it is aware of *X*?

The search for the functions of consciousness has often led to dead ends. For instance, perceptual responses, linguistic analysis, number processing and even executive functions may be triggered non-consciously (Van Gaal et al., 2012), and the famous Libet experiments and phenomena such as choice blindness imply that consciousness is merely a post-hoc rationalizer of what we do or say (Johansson et al., 2005; Libet et al., 1983). But just because this search has ruled out some promising candidates for the functions of consciousness does not mean there are none to be found. As Doerig et al. conclude: 'Maybe consciousness is a "solution", a by-product, or a core component of a computational challenge that information processing systems need to solve – and that we have not discovered yet.'

One candidate function is that consciousness is for sharing (Frith, 2008). Unconscious mental states might be sufficient to get by on our own. But sharing mental states with others often requires conscious report. Consider two hunters stalking a deer. They might wish to share information about what they are each seeing, both to pool their perceptual resources and to provide

CONTACT Stephen M. Fleming stephen.fleming@ucl.ac.uk Wellcome Centre for Human Neuroimaging and Department of Experimental Psychology, University College London, London, UK

'common ground' for future coordinated action (Roepstorff & Frith, 2004). Experiments show that when two individuals share their confidence in their percepts, they can arrive at a joint decision that is better than that of the best individual working alone – a form of interpersonal Bayesian cue-combination (Bahrami et al., 2010).

Identifying candidate functions for consciousness can constrain theorizing on its algorithms and neural substrates. For instance, for awareness reports to be socially useful they need to be flexible (we can communicate awareness of a deer, a lion, or some berries with equal fluency) and simple (occupying an abstract, low-dimensional space ranging from unaware to aware). Accordingly, linguistic analysis suggests that pairs of naïve observers rapidly hit upon a common confidence scale with which to communicate what they are seeing in psychophysical experiments (Fusaroli et al., 2012). Recent proposals including higher-order state space (HOSS) and predictive global neuronal workspace (PGNW) models have begun to flesh out how high-dimensional (but potentially unconscious) perceptual representations 'connect up' to the capacity for such reports (Bengio, 2017; Fleming, 2020; Whyte & Smith, 2020). We can also make informed guesses about when the functionality for sharing mental states became useful during evolution. For instance, computer simulations show that the enhanced visual range of terrestrial compared to aquatic animals may have driven the need for planning based on the perception of distant objects (Mugan & MacIver, 2020). Finally, functional criteria that constrain algorithms for awareness can help in interpreting candidate neural substrates (the implementational level). For instance, neural recordings in primates indicate that the dorsolateral prefrontal cortex (area 46) contains cells that show stimulus-independent representation of stimulus presence and absence – consistent with encoding a low-dimensional axis of awareness (Merten & Nieder, 2012; Passingham & Lau, 2019).

As Doerig et al. highlight, a search for algorithms that are unmoored from function (such as causal structure theories) is liable to lead to an unconstrained exploration of often beautiful but potentially meaningless mathematics. A renewed focus on function instead aligns more naturally with recent illusionist approaches to consciousness (Frankish, 2016; Graziano, 2016). Conscious awareness can sometimes seem like magic, appearing from nowhere out of a physical system. But a scientific explanation of a magic show cannot be given in terms of magic itself, but in terms of why the audience believe that magic is a plausible explanation of a trick, and why they go on to amazedly tell their friends about it. In the same way, rather than fret about the mystery of intrinsic subjectivity, the illusionist attempts to explain what people believe (and can communicate) about their experience using the lingua franca of cognitive science – a science of both the trick and the audience. Theories of consciousness constrained by social function can go one step further, and explain why putting on the show might have been useful in the first place.

Disclosure statement

No potential conflict of interest was reported by the author.

Funding

This work was supported by a Philip Leverhulme Prize from the Leverhulme Trust and a Wellcome/Royal Society Sir Henry Dale Fellowship (206648/Z/17/Z).

References

Bahrami, B., Olsen, K., Latham, P. E., Roepstorff, A., Rees, G., & Frith, C. D. (2010). Optimally interacting minds. *Science*, *329* (5995), 1081–1085. https://doi.org/10.1126/science.1185718

Bengio, Y. 2017. The consciousness prior *arXiv*, arXiv:1709.08568v2

Fleming, S. M. (2020). Awareness as inference in a higher-order state space. *Neuroscience of Consciousness*, *2020*(1), niz020. https://doi.org/10.1093/nc/niz020

Frankish, K. (2016). Illusionism as a theory of consciousness. *Journal of Consciousness Studies*, *23*(11–12), 11–39. https://www.ingentaconnect.com/content/imp/jcs/2016/00000023/f0020011/art00002

Frith, C. D. (2008). The Social functions of consciousness. In L. Weiskrantz & M. Davies (Eds.), *Frontiers of consciousness: Chichele lectures* (pp. 225–244). Oxford University Press.

Fusaroli, R., Bahrami, B., Olsen, K., Roepstorff, A., Rees, G., Frith, C., & Kristian, T. (2012). Coming to terms: Quantifying the benefits of linguistic coordination. *Psychological Science*, *23*(8), 931–939. https://doi.org/10.1177/0956797612436816

Graziano, M. S. A. (2016). Consciousness engineered. *Journal of Consciousness Studies*, *23*(11–12), 98–115. https://www.ingentaconnect.com/content/imp/jcs/2016/00000023/f0020011/art00008

Johansson, P., Hall, L., Sikström, S., & Olsson, A. (2005). Failure to detect mismatches between intention and outcome in a simple decision task. *Science*, *310*(5745), 116–119. https://doi.org/10.1126/science.1111709

Libet, B., Wright, E. W., Jr, & Gleason, C. A. (1983). Preparation-or intention-to-act, in relation to pre-event potentials recorded at the vertex. *Electroencephalography and Clinical Neurophysiology*, *56*(4), 367–372. https://doi.org/10.1016/0013-4694(83)90262-6

Marr, D., & Poggio, T. (1976). From understanding computation to understanding neural circuitry. *Massachusetts Institute of Technology Artificial Intelligence Laboratory*, *A.I. Memo*, *357*. http://hdl.handle.net/1721.1/5782

Merten, K., & Nieder, A. (2012). Active encoding of decisions about stimulus absence in primate prefrontal cortex neurons. *Proceedings of the National Academy of Sciences*,

109(16), 6289–6294. https://doi.org/10.1073/pnas.1121084109

Mugan, U., & MacIver, M. A. (2020). Spatial planning with long visual range benefits escape from visual predators in complex naturalistic environments. *Nature Communications*, *11* (1), 1–14. https://doi.org/10.1038/s41467-020-16102-1

Passingham, R. E., & Lau, H. C. (2019). Acting, seeing, and conscious awareness. *Neuropsychologia*, *128*, 241–248. https://doi.org/10.1016/j.neuropsychologia.2017.06.012

Roepstorff, A., & Frith, C. (2004). What's at the top in the top-down control of action? Script-sharing and 'top-top'control of action in cognitive experiments. *Psychological Research*, *68*(2–3), 189–198. https://doi.org/10.1007/s00426-003-0155-4

Van Gaal, S., De Lange, F. P., & Cohen, M. X. (2012). The role of consciousness in cognitive control and decision making. *Frontiers in Human Neuroscience*, *6*, 121. https://doi.org/10.3389/fnhum.2012.00121

Whyte, C. J., & Smith, R. (2020). The predictive global neuronal workspace: A formal active inference model of visual consciousness. *Progress in Neurobiology*, 101918. https://doi.org/10.1016/j.pneurobio.2020.101918