



Reviewed by  
**IAN PHILLIPS &  
SIMON A. B. BROWN**

[Home](#)

## IN CONSCIOUSNESS WE TRUST

Hakwan Lau

Reviewed by Ian Phillips and Simon A B Brown

*In Consciousness We Trust: The Cognitive Neuroscience of Subjective Experience*

Hakwan Lau

Oxford: Oxford University Press, 2022, £24.99

ISBN 9780198856771

Hakwan Lau is well known as a unique and forceful voice in the science of consciousness. His first book, *In Consciousness We Trust* offers a characteristically original and opinionated treatment of the field. Unusually for a book by a neuroscientist, it draws on and engages extensively with contemporary philosophy. Unusually for an academic monograph, it is a striking admixture of memoir and manifesto. The result is both appealing and accessible. It will be of wide interest to neuroscientists, philosophers, sociologists of science, and simply curious outsiders.

Lau sees a field at a crossroads, where it could either prioritize 'getting the empirical facts right' (p. 8) and enter the scientific mainstream, or degenerate into an 'increasingly esoteric and theoretically indulgent' backwater (p. 10). In moments of polemic, Lau rails against his discipline's 'unduly heavy focus on personal glory and stardom' (p. 8) and reliance on 'a few wealthy private donors' (p. 10). Indeed, reflecting on the socio-structural reasons behind this culture, Lau ends the book by suggesting that reforming the field's politics is 'the Truly Hard Problem of consciousness' (p. 217). But the central goal of the book is to make intellectual progress, a goal primarily pursued by mapping the landscape, assessing current views, and ultimately advocating for his own distinctive 'perceptual reality monitoring' account.

We reach this account in part through memoir, as Lau recounts his journey—and that of the field’s—studying the elusive and sometimes forsaken quarry of consciousness. Throughout, Lau pays tribute to the many brilliant scientists he has worked with, first as mentee (notably, of the distinguished British neuroscientist Dick Passingham) and then as (clearly, deeply committed) mentor. He also reflects with rare candour on past mistakes, acknowledging flaws in earlier studies, failures of replication, and difficulties with interpretation. Lau’s positive view, however, is presented as manifesto—‘A Centrist Manifesto’, to use the title of Chapter 6. To understand Lau’s platform, we need first to consider his map of consciousness science as a whole.

Not unusually (for example, Block [2019]), Lau views (‘serious’, ‘mainstream’) consciousness science as comprising two broad camps. For global theorists (for example, Baars [1988]; Dehaene and Naccache [2001]), representations are conscious just when they are made available to a wide range of cognitive systems such as attention, memory, planning, and report. Such global ‘broadcast’ is widely held to be implemented by a ‘central workspace’, constituted by long-range neurons predominantly in prefrontal and parietal cortices. For local theorists (for example, Lamme [2010]; Block [2011]), certain sensory activities in occipital and temporal cortices suffice for perceptual consciousness. Consciousness neither requires recruitment into the workspace, nor wider cognitive availability.

In the first half of his book, Lau offers an expansive critique of both camps, before proposing his own ‘moderate, ‘centrist’ position’ (p. 139) between these ‘polar extremes’ (p. 25). We turn to these critiques shortly. But first, two comments are worth making about this framing.

The first concerns what is not on Lau’s map. Seth and Bayne ([2022]) list a ‘selection’ of twenty-two theories of consciousness, many of which find no obvious home in Lau’s landscape. At least some of these, like integrated information theory (Tononi et al. [2016]), Lau hopes to exclude from readers’ conception of the scientific mainstream, considering them examples of the theoretically indulgent backwaters mentioned above (p. 7)—arguably with some justification (Bayne [2018]). But another important family of theories missing from even Seth and Bayne’s selection, which we think deserves to be taken more seriously, adopts an evolutionary biological perspective. These theories often emphasize non-cortical circuits whose function is to integrate sensory information with information concerning an organism’s homeostatic needs into an agent-environment model (Merker [2007]; Barron and Klein [2016]; cf. Mallatt and Feinberg [2018]; Ginsburg and Jablonka [2019]; Godfrey-Smith [2020]).

The second comment concerns a background assumption of Lau’s, and indeed the field’s as a whole, that methodological challenges to a science of consciousness can be overcome. Lau is well aware of such challenges, providing an extremely valuable catalogue of confounds that he sees as plaguing many published experiments. Suppose we attempt to isolate the neural circuits correlated with conscious experience by contrasting activity between a conscious and an unconscious condition. How can we be sure that the difference in activity exclusively reflects the difference in consciousness between the conditions as opposed to many other confounding factors, such as the fact that different stimuli were used in the two conditions or the fact that the subject reported—or thought—that they saw something in one condition but not the other? Controversially, Lau sees the most serious, and least addressed, confound as that of task-performance capacity. His concern is that the difference between conditions might reflect not consciousness, but differences in signal available to higher-level areas and attendant differences in task-performance potential. As Lau acknowledges, this is tendentious. Some will think that task-performance capacity is constitutive of consciousness rather than a confounding factor, just as we would not treat two individuals’ tendencies to be publicly recognized as a confounder in comparing how famous they were. Also tendentious is Lau’s optimism that such confounds can be surmounted, leaving us with a familiar project of testing and choosing between rival theories. An alternative perspective perceives a more intractable methodological challenge to the very possibility

of a science of consciousness than Lau allows (cf. Irvine [2013]; Phillips [2018]). From this perspective, the problematic status of conscious science is traceable to intellectual and not merely socio-cultural sources.

These comments noted, let us return to Lau's landscape and his critique of its two dominant parties.

### Problems for global theories

Against global theories, Lau objects that while frontal activity is critical for conscious experience, its involvement is modest. Consciousness does not require the extensive activity that global theories predict (p. 129). However, many of his criticisms run in the other direction, claiming that surprisingly sophisticated processing occurs without awareness, including establishment of a 'task set' (that is, the activation of an appropriate set of cognitive processes for a given task; see, for example, Lau and Passingham [2007]), cognitive control (see, for example, van Gaal *et al.* [2010]), and working memory (see, for example, Trübtschek *et al.* [2017]).

Lau frankly recognizes that such studies fail to establish the complete absence of awareness, either because they rely on potentially biased subjective measures (such as asking a subject whether or not they saw anything—a measure that risks underestimating awareness if the subject is cautious in saying 'yes'), or because of insufficient statistical power. Nonetheless, Lau does deem 'a few findings [...] relatively promising' (p. 115), highlighting two studies of patient GY who has 'blindsight', a condition widely held to involve extensive unconscious processing. In the first, Persaud and Cowey ([2008]) demonstrated cases where GY performed above-chance in reporting the interval in which a stimulus was located in his blindfield despite being instructed to report where it was not located. In the second, GY's choices in a betting task suggested inferior metacognitive sensitivity for responses made to stimuli in his blindfield (Persaud *et al.* [2011]). It is unclear, however, why Lau does not see the same issues about bias and weak awareness as arising in these cases (see Phillips [2021a], and the subsequent exchange in Michel and Lau [2021]; Phillips [2021b]).

Lau also mentions aphantasia—reported lack of conscious imagery—as a potential counter-example to global theories, as such reports can co-exist with high performance in working-memory tasks. But many unresolved issues arise, including how to objectively measure true absence of imagery experience (though see Kay *et al.* [2022]), and whether the bases of performance in working memory tasks are unconscious visual representations, or instead conscious spatial and/or amodal imagery or non-imagistic representations. Another tantalizing suggested line of evidence is that strong non-conscious effects may be produced via decoded neuro feedback (pp. 119–22). This exciting, though still nascent, technique—which Lau and his colleagues have helped pioneer—exploits real-time fMRI pattern analysis to pair non-conscious neural signals with reward or punishment to generate long-term conditioned responses. For example, subjects are given a reward when their brain activity matches the patterns exhibited during presentation of a stimulus with negative associations (for example, a snake). Even though, when given feedback, subjects are not presented with a snake, are not consciously thinking about snakes, and are unaware that this is what their brain activity corresponds to, the feedback still reduces their negative physiological responses when next shown a snake.

In addition to empirical objections, Lau emphasizes two considerations more familiar to philosophers, namely, that global views cannot account for the apparent richness of perceptual experience and that they predict (in Lau's view implausibly) 'that very simple computer programs and robots may be conscious' (p. 131). These ideas, to which we return below, reflect two core manifesto commitments: to take subjective reports seriously and to avoid overly liberal attributions of consciousness. Blindsight being unconscious represents a related third commitment, for it is a case

where we ought to take subjective denials of experience seriously and which provides a key ground for thinking that substantial cognition can occur outside consciousness.

### **Problems for local theories**

Against local theories, Lau points to a mismatch problem, claiming that there are many cases in which early sensory activity does not correspond to conscious content. One example is binocular rivalry, where conscious experience alternates between two different stimuli, one presented to each eye. In rivalry, Lau claims, early sensory activity is equally strong for both stimuli, and we need to look to prefrontal areas to find correlates of conscious experience (pp. 41–44). Lau also points to the stability of our conscious experience, contrasting this with the constant modulation of sensory activity by attention (pp. 88–89). Finally, Lau mentions that blindsight patients lacking primary visual cortex (V1) nonetheless have conscious motion experiences (p. 134).

These are important considerations. However, they primarily threaten simpler local theories that propose that our conscious experience directly reflects any and all V1 activity. But local theorists need not focus exclusively on V1 (for example, Block [2007], p. 499, Footnote 10). Nor need they claim that all activity in sensory areas corresponds to conscious experience. For example, Lamme ([2015]) specifically associates consciousness with recurrent activations, treating purely feedforward activity as non-conscious (though for a critique of Lamme's specific position, see p. 134).

Lau also points to complaints more familiar to philosophers, namely, that local theories are anti-functionalist and (Lau alleges) risk falling into panpsychism (pp. 136–37, 144). Panpsychism falls foul of Lau's pledge to avoid overly liberal attributions of consciousness—if simple robots aren't conscious, then electrons certainly aren't! Functionalism represents a further critical commitment for Lau.

For Lau, functionalism is the view that 'information processing is all that matters' (p. 138), a thesis he calls 'the very premise of cognitive neuroscience' (p. 137). Functionalism, thus construed, demands that accounts of subjective experience be substrate independent. Lau assumes that localists must deny this. However, while some localists do espouse a theory of consciousness on which biological instantiation matters, this is not obligatory. Nothing prevents a localist from holding that the information processing performed in sensory areas, characterized in a substrate-independent fashion, is the basis of conscious experience, provided the processing is sufficiently complex and distinctive to avoid implausible attributions of consciousness. Substrate independence is also not obvious. Those who claim that biological instantiation matters press that we are far from having a complete theory of the mind and would do better to adopt a bottom-up and evolutionary perspective, charting the natural emergence of simple forms and precursors of consciousness, as opposed to a one-size-fits-all conception of mind as computer (again, see Merker [2007] and references above).

Lau's criticisms set the stage for his positive proposal, which he argues gives us the best of both global and local views. This theory will predict modest frontal activity, be thoroughly functionalist, avoid overly liberal attributions of consciousness, and, finally, take seriously subjective introspective reports of richness, stability, and, in the case of blindsight, absence of experience. Enter perceptual reality monitoring.

### **Lau's account of consciousness**

Lau postulates a prefrontal perceptual reality monitoring (PRM) mechanism, which functions to determine the status of activity in sensory areas—roughly, whether it is noise or driven by sensory stimulation. It does this automatically and implicitly, as opposed to relying on introspection and explicit reasoning. The result is a de-intellectualized higher-

order theory. Specifically, Lau proposes that PRM mechanisms exploit computations resembling a discriminator in a generative adversarial network, a computational architecture that he thinks recent work in machine learning independently suggests we should expect to find in visual processing (cf. Gershman [2019]). The PRM mechanism then 'tags' sensory activity: being tagged as perception rather than noise renders a first-order state phenomenally conscious. Phenomenal consciousness has numerous downstream effects, as the tag is used by other systems to determine how to use the first-order state. One effect Lau emphasizes is that representations tagged as perception gain 'assertoric force': a tendency to be believed, which needs to be actively inhibited if we otherwise know they are false, as with stubborn illusions and magic tricks. Such tagging also causally facilitates (although does not entail) metacognition, inhibition, cognitive control, and global broadcast. Thus, consciousness is 'the "gating" mechanism through which perception impinges on higher cognition' (p. 178).

How does this view accommodate the desiderata identified above?

First, it predicts the modest frontal activity Lau thinks is associated with consciousness: With local views, Lau's PRM theory sees low-level sensory activity as constitutively contributing to perceptual consciousness, and allows that phenomenal consciousness occurs without global broadcast. But with global views, it denies that sensory activity is sufficient for phenomenal consciousness, insisting on some prefrontal activity—the activity instantiating reality monitoring.

Second, the theory is thoroughly functionalist: reality monitoring is describable in entirely computational, substrate-independent terms, as Lau's appeal to generative adversarial networks may help illustrate.

Third, since perceptual reality monitoring is not broadly instantiated, Lau avoids very liberal attributions of consciousness. That said, Lau's position on exactly which systems are likely to be phenomenally conscious is unusual: he expresses measured scepticism about many non-human species, while accepting that though most simple robots are not conscious, we could build relatively simple ones which were (pp. 166–69). Lau's position on animals also raises unanswered questions: how is it that animals without a prefrontal cortex (and, for Lau, a PRM mechanism) are able to operate effectively, given that they presumably have just as much noise in their sensory areas as humans? If they have some other mechanism for filtering noise, is that mechanism associated with consciousness too, and why do humans use PRM?

Fourth, Lau can say that blindsight patients' perception is unconscious, in line with their reports, by claiming that it involves untagged but still efficacious sensory activity.

Fifth and lastly, Lau devotes a whole chapter to phenomenal richness. It is widely held that only a relatively meagre subset of sensory representations is globally broadcast. Consequently, local theorists contend that only they can account for the relative richness of experience (Block [2011]). In reply, global theorists typically deny that experience is as rich as it seems, citing paradigms such as change and inattention blindness (Cohen *et al.* [2016]). For Lau, neither view is satisfactory. Subjects do report richness and we should take their reports seriously, albeit not as infallible (p. 92). Yet they also report stable phenomenology, and local sensory activity is in constant flux (p. 102).

Lau's middle way on richness appeals to 'inflation'. Consider the periphery, where subjects report rich experience despite an impoverished signal. According to inflation, subjects think that they see many items because (signal detection theoretic analyses suggest) they adopt more liberal criteria than in central vision, and so weaker sensory signals suffice to trigger an estimate that a stimulus is present (Odegaard *et al.* [2018]). Lau's discussion of these

results admits two quite different interpretations, however. On one interpretation, our experience is impoverished, but we mistakenly judge it rich due to applying liberal criteria to unconscious signals. On the other interpretation, liberal criteria affect perceptual reality monitoring itself, so that weaker signals can trigger a 'genuine perception' tag for peripheral activity. On this interpretation, consciousness is rich, representing much more content than is represented by globally broadcast states. It is just that such representations are unreliable. This latter interpretation is a particularly interesting addition to the richness debate, offering as it does a non-localist explanation of genuine phenomenal richness.

### **What it's like**

Lau is admirably hesitant to claim to have solved the hard problem of consciousness—perhaps reasonably so, given that it seems easy enough to conceive of 'zombie' systems with PRM mechanisms. Still, we might hope for a theory that captures or at least connects to some of the traditionally puzzling features of phenomenal consciousness. For example, it is reasonable to hope for a theory that in some way sheds light on the ideas that consciousness implies a subject, that it has a first-personal or perspectival nature, and that there is something it is like to be in a conscious state—such as seeing a red rose.

Lau does think that we can partially explain why there is something it is like to see red by appeal to Rosenthal's quality-space theory. Rosenthal ([2010])'s theory of mental qualities is based around constructing theoretical spaces of stimuli, in which the distance between stimuli reflects their discriminability for the individual in question. On this theory, mental qualities are the properties by virtue of which perceptual discriminations are made, so there will be matching spaces for mental qualities, with distance corresponding to similarity in experience. This neatly explains why we can articulate some comparative facts about our experiences, such as that seeing red is more like seeing orange than seeing green, and completely unlike hearing a flute.

Lau thinks quality-space theory both captures an aspect of what it is like to see red, and coheres nicely with PRM, which he thinks independently motivates a special role for quality-spaces. At root, this role derives from downstream systems needing to be able to interpret which activity has been tagged by PRM mechanisms in order to treat it appropriately. A natural way to achieve this would involve the tag's including a hyperlink-like pointer to the relevant activity. In sensory cortices, we find spatially similar (neurally overlapping) representations for similar stimuli. This, Lau thinks, means that pointers could systematically and compactly refer to sensory activity in terms of its location in the space of possible sensory activity, without reproducing its contents in detail. If PRM mechanisms do represent first-order states in this way, Lau suggests, they would implicitly 'know' about the stimulus-space for free: similar states will be indexed with similar addresses. Moreover, states that are similarly addressed will be harder to discriminate, hence this implicitly known stimulus space will correspond to quality space. For Lau, the implicit representation of quality spaces by the same mechanisms responsible for consciousness would at least partially account for why subjective experience can seem hard to articulate: we do not typically have detailed higher-order representations of the contents of subjective experiences, just pointers to sensory representations, which in turn are in an analogue format that does not translate straightforwardly into natural language.

### **Extensions of the PRM theory and future directions**

Lau's positive theory draws together a number of ideas in a valuable and interesting way, but remains in many respects underspecified; *In Consciousness We Trust* is best thought of as a roadmap for further work supported by a promising sketch of a view rather than a full-blown theory. For example, Lau says little at a detailed computational level about how PRM operates (beyond a broad analogy to generative adversarial networks), or at a detailed

neurobiological level (beyond locating the PRM mechanism in the prefrontal cortex). It is also unclear exactly what PRM mechanisms tag: is any activity in sensory areas fair game, from the activity of individual neurons to population-level activity for subsystems, particular levels of the processing hierarchy, or the entire processing hierarchy as a whole?

Lau's theory also awaits specification in its application to different kinds of conscious state. The overwhelming focus of Lau's book is perceptual, and especially visual, consciousness. He does offer proposals about other conscious states, including other perceptual modalities, emotions, memory, dreaming, and volition (see further Lau *et al.* [2022]), but such proposals are more tentative and raise various complications.

Even the extension beyond vision to senses like olfaction is not entirely straightforward. The basic structure of his account—a reality monitor determining the status of activity in sensory areas—might apply. Yet olfactory areas are not neatly organized spatially in the way that visual areas are, so PRM may need to address such activity using a different kind of pointer with a looser connection to quality spaces. Lau claims such areas are '*functionally* spatially analog [...] That is to say, let's say we hypothetically scramble around the neurons in the visual system of a mammal, while keeping the connections between the neurons the same. This should not fundamentally change the basic computational properties of the circuit' (p. 207). It is unclear to us that this notion does the work Lau hopes; certainly, more detail is needed.

The extension to sensory memory and imagery raises further questions. Rather than simply distinguishing between sensory activity reflecting noise and external signal, the reality monitor must also consider different potential kinds of endogenously generated sensory activity and tag such activity appropriately as memory, counterfactual imagination, and so on. (pp. 159–60). But does a single decision, located in one system, determine the status of sensory activity as noise, perception, memory, imagination, or still yet another category? Or is there something more like a canonical computation that makes related determinations for different states and purposes? Lau also suggests that additional resources will be needed to capture aspects of the phenomenology of memory related to the self (p. 209).

Lau admits that his account is also incomplete for other cognitive states, such as thinking and emotions, that involve processing rather different to the sort of sensory activity perceptual reality monitoring monitors (pp. 165, 183–86; LeDoux and Lau [2020]). Lau has much to say about cognition in Chapter 8, making wide-ranging (albeit extremely speculative) proposals about the role of a 'narrative system' that has access to perception as vetted by the PRM, even attempting to connect consciousness as studied in his field to notions like the Marxist concept of 'false consciousness' (p. 179). However, he does not offer a detailed account of the phenomenology of cognitive states themselves—the difference between what it is like to think about calculus and cricket, for example. While some have claimed that all consciousness is perceptual (Prinz [2007]), this is far from a universal view. If there is non-perceptual phenomenology, it is obscure not only how the quality space view could apply to such states, but also what reality monitoring would look like. Successfully extending Lau's framework to these cases would constitute a major advance.

## Conclusions

As a memoir written midway through Lau's career, *In Consciousness We Trust* offers an opportunity to look forward as well as back. Indeed, Lau has just begun a new phase of life as leader of a new consciousness lab at the RIKEN Center for Brain Science. What can we expect the future of consciousness science to hold?

If Lau's ambitions are realized, we will see a reconfiguration of both the intellectual and sociocultural landscape. Sociologically, consciousness science will become part of mature, mainstream science. Researchers will develop ever-more sophisticated computational models and use agreed-on experimental paradigms to circumvent confounds. Funding will come from major agencies—an ambition being pursued vigorously by Lau's former mentees Megan Peters and Brian Odegaard—and less hostage to the fads of celebrities, philosophers, and the media. Intellectually, the old two-party duopoly of globalist and localist theories will be disrupted by Lau's third way offering—a theory that promises that higher-order theories need not be over-intellectualist and that prefrontal activation requirements are consistent with subjective richness, and which has surprising consequences (for example, that while relatively simple robots equipped with the right kind of PRM mechanisms might be conscious, other robots and many non-human animals are likely not).

Intellectual fortunes are no easier to forecast than political ones. But it is worth recognizing that Lau reconfigures two-party politics from within a worldview where these parties exhaust the options, a reconfiguration some will find more reactionary than revolutionary. *Pace* Lau, more radical positions, including views that reject top-down computationalist functionalism, are not limited to panpsychism and integrated information theory. As noted, we might instead turn to bottom-up approaches to consciousness founded in biology. We might also see the concerns and confounds affecting both past and present work as the sign of a deeper methodological challenge to a science of consciousness than Lau allows. But wherever one places one's bets, Lau's book is a must-read for anyone interested in the practice and promise of contemporary consciousness science.

### Acknowledgments

Thanks to Howard Egeth, Chris Fetsch, Steven Gross, Jorge Morales, Caroline Myers, and Kia Torab for extremely helpful comments on an earlier draft.

*Ian Phillips*  
Johns Hopkins University  
[ianbphillips@jhu.edu](mailto:ianbphillips@jhu.edu)

*Simon A. B. Brown*  
Johns Hopkins University  
[sbrow285@jhu.edu](mailto:sbrow285@jhu.edu)

### References

Baars, B. J. [1988]: *A Cognitive Theory of Consciousness*, Cambridge: Cambridge University Press.

Barron, A. B. and Klein, C. [2016]: 'What Insects Can Tell Us about the Origins of Consciousness', *Proceedings of the National Academy of Sciences USA*, **113**, pp. 4900–8.

Bayne, T. [2018]: 'On the Axiomatic Foundations of the Integrated Information Theory of Consciousness', *Neuroscience of Consciousness*, **2018**, available at



- Block, N. [2007]: 'Consciousness, Accessibility and the Mesh between Psychology and Neuroscience', *Behavioral and Brain Sciences*, **30**, pp. 481–548.
- Block, N. [2011]: 'Perceptual Consciousness Overflows Cognitive Access', *Trends in Cognitive Sciences*, **15**, pp. 567–75.
- Block, N. [2019]: 'What Is Wrong with the No-Report Paradigm and How to Fix It', *Trends in Cognitive Sciences*, **23**, pp. 1003–13.
- Cohen, M. A., Dennett, D. C. and Kanwisher, N. [2016]: 'What Is the Bandwidth of Perceptual Experience?', *Trends in Cognitive Sciences*, **20**, pp. 324–35.
- Debner, J. A. and Jacoby, L. L. [1994]: 'Unconscious Perception: Attention, Awareness, and Control', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **20**, pp. 304–17.
- Dehaene, S. and Naccache, L. [2001]: 'Towards a Cognitive Neuroscience of Consciousness: Basic Evidence and a Workspace Framework', *Cognition*, **79**, pp. 1–37.
- Gershman, S. J. [2019]: 'The Generative Adversarial Brain', *Frontiers in Artificial Intelligence*, **2**, p. 18.
- Ginsburg, S. and Jablonka, E. [2019]: *The Evolution of the Sensitive Soul: Learning and the Origins of Consciousness*, Cambridge, MA: MIT Press.
- Godfrey-Smith, P. [2020]: *Metazoa: Animal Life and the Birth of the Mind*, New York: Harper-Collins.
- Irvine, E. [2013]: *Consciousness as a Scientific Concept*, Dordrecht: Springer.
- Kay, L., Keogh, R., Andriillon, T. and Pearson, J. [2022]: 'The Pupillary Light Response as a Physiological Index of Aphantasia, Sensory, and Phenomenological Imagery Strength', *eLife*, **11**, available at .
- Lamme, V. A. F. [2010]: 'How Neuroscience Will Change Our View on Consciousness', *Cognitive Neuroscience*, **1**, pp. 204–20.
- Lamme, V. A. F. [2015]: *The Crack of Dawn: Perceptual Functions and Neural Mechanisms that Mark the Transition from Unconscious Processing to Conscious Vision*, Mainz: Johannes Gutenberg-Universität Mainz.
- Lau, H. C. and Passingham, R. E. [2007]: 'Unconscious Activation of the Cognitive Control System in the Human Prefrontal Cortex', *The Journal of Neuroscience*, **27**, pp. 5805–11.
- Lau, H., Michel, M., LeDoux, J. E. and Fleming, S. M. [2022]: 'The Mnemonic Basis of Subjective Experience', *Nature Reviews Psychology*, **2022**, available at .
- LeDoux, J. E. and Lau, H. [2020]: 'Seeing Consciousness through the Lens of Memory', *Current Biology*, **30**, pp. 1018–22.

Mallatt, T. E. and Feinberg, J. M. [2018]: *Consciousness Demystified*, Cambridge, MA: MIT Press.

Merker, B. [2007]: 'Consciousness without a Cerebral Cortex: A Challenge for Neuroscience and Medicine', *Behavioral and Brain Sciences*, **30**, pp. 63–81.

Michel, M. and Lau, H. [2021]: 'Is Blindsight Possible under Signal Detection Theory? Comment on Phillips (2021)', *Psychological Review*, **128**, pp. 585–91.

Odegaard, B., Chang, M. Y., Lau, H. and Cheung, S.-H. [2018]: 'Inflation versus Filling-In: Why We Feel We See More Than We Actually Do in Peripheral Vision', *Philosophical Transactions of the Royal Society B*, **373**, available at .

Persaud, N. and Cowey, A. [2008]: 'Blindsight is Unlike Normal Conscious Vision: Evidence from an Exclusion Task', *Consciousness and Cognition*, **17**, pp. 1050–55.

Persaud, N., Davidson, M., Maniscalco, B., Mobbs, D., Passingham, R. E., Cowey, A. and Lau, H. [2011]: 'Awareness-Related Activity in Prefrontal and Parietal Cortices in Blindsight Reflects More Than Superior Visual Performance', *NeuroImage*, **58**, pp. 605–11.

Phillips, I. B. [2018]: 'The Methodological Puzzle of Phenomenal Consciousness', *Philosophical Transactions of the Royal Society B*, **373**, pp. 1–9.

Phillips, I. B. [2021a]: 'Blindsight Is Qualitatively Degraded Conscious Vision', *Psychological Review*, **128**, pp. 558–84.

Phillips, I. B. [2021b]: 'Bias and Blindsight: A Reply to Michel and Lau', *Psychological Review*, **128**, pp. 592–95.

Prinz, J. [2007]: 'All Consciousness Is Perceptual', in B. P. McLaughlin and J. D. Cohen (eds), *Contemporary Debates in Philosophy of Mind*, Oxford: Blackwell, pp. 335–57.

Rosenthal, D. [2010]: 'How to Think about Mental Qualities', *Philosophical Issues*, **20**, pp. 368– 93.

Seth, A. K. and Bayne, T. [2022]: 'Theories of Consciousness', *Nature Reviews Neuroscience*, **23**, pp. 439–52.

Tononi, G., Boly, M., Massimini, M. and Koch, C. [2016]: 'Integrated Information Theory: From Consciousness to Its Physical Substrate', *Nature Reviews Neuroscience*, **17**, pp. 450–61.

Trübtschek, D., Marti, S., Ojeda, A., King, J.-R., Mi, Y., Tsodyks, M. and Dehaene, S. [2017]: 'A Theory of Working Memory without Consciousness or Sustained Activity', *eLife*, **6**, available at .

van Gaal, S., Ridderinkhof, K. R., Scholte, H. S. and Lamme, V. A. [2010]: 'Unconscious Activation of the Prefrontal No-Go Network', *Journal of Neuroscience*, **30**, pp. 4143–50.

f

🐦