Journal Club

Editor's Note: These short reviews of recent *JNeurosci* articles, written exclusively by students or postdoctoral fellows, summarize the important findings of the paper and provide additional insight and commentary. If the authors of the highlighted article have written a response to the Journal Club, the response can be found by viewing the Journal Club at www.jneurosci.org. For more information on the format, review process, and purpose of Journal Club articles, please see https://www.jneurosci.org/content/jneurosci-journal-club.

Reconciling Current Theories of Consciousness

Sébastien Maillé* and Michael Lynn*

Neuroscience Graduate Program, Department of Cellular and Molecular Medicine, University of Ottawa, Ottawa, Ontario K1G 8M5, Canada

Review of Noel et al.

Understanding the neural basis of consciousness is one of the fundamental challenges in modern neuroscience. A number of sophisticated models and theories have attempted to formalize how the brain implements consciousness using insights from philosophy, psychology, computer science, and neuroscience. These include two major and perhaps competing theories, the integrated information theory (IIT) and the global neuronal workspace (GNW) theory, which differ mainly in their level of conceptual abstraction and anatomical specificity.

The IIT, first proposed by Tononi (2004), focuses on defining what a conscious system should look like with respect to information processing and architecture without considering particular brain areas or temporal profiles. One prediction of IIT is that neural networks supporting consciousness must be highly interconnected, effectively integrating different components of a state into a unified experience. A crucial advantage of the IIT is that it provides a mathematical metric of irreducibility (or integration), Φ, that can be related to the level of consciousness. Proponents of IIT point to its explanatory power: for instance, it can explain why the cortex is capable of producing conscious experience while the cerebellum is not (Lemon and Edgley, 2010; Yu et al., 2015), even though the cerebellum possesses up to four times more neurons. While the IIT has not received unambiguous validation (possibly due to the abstract nature of its description of consciousness; for review, see Tononi et al. (2016)), it provides one of the most detailed accounts for the emergence of conscious experience from an information-processing network.

The GNW theory (Dehaene and Changeux, 2011), in contrast to the IIT, was empirically derived from EEG and imaging studies in humans and primates. These studies have shown that when a stimulus is presented but not consciously perceived, activation can be seen mainly in the associated primary sensory cortices. When the stimulus is consciously perceived, however, activation in primary cortical areas is followed by a delayed "neural ignition" in which a sustained wave of activity propagates across prefrontal and parietal association cortices. According to the GNW model, this allows relevant information to be broadcast across the brain to other subsystems for use in decision-making, reporting, memory consolidation, and other processes. Thus, while IIT focuses on abstract connectivity and information-processing structure, GNW proposes a concrete spatiotemporal locus for conscious processes.

Unfortunately, while both IIT and GNW have obtained experimental support, testable predictions from both theories are seldom compared within the same dataset. In a recent issue of *The Journal of Neuroscience*, Noel et al. (2019) leveraged a previously published experimental dataset to directly compare IIT and GNW at the single-unit level. In the experiments published by Ishizawa et al. (2016) (and later reanalyzed by Noel et al., 2019), monkeys were subjected to nonaversive stimuli while extracellular microelectrode arrays recorded single-unit activity in S1 (primary somatosensory cortex) and vPM (ventral premotor cortex; involved in multisensory integration). The stimuli consisted of an auditory stimulus, a tactile stimulus, or concurrent auditory and tactile stimuli. Crucially, part of the way through the task monkeys were anesthetized with propofol, permitting a sophisticated comparison of single-neuron activity across states of consciousness.

The authors exploited the multisensory nature of the stimuli to categorize neurons based on information processing rules, which they could relate to key mathematical predictions from IIT. Neurons were classified as integrative (AND gate; exhibiting a multisensory response greater than the largest unisensory response) or convergent (XOR gate; exhibiting a multisensory response smaller than or equal to the largest unisensory response). Ac-

Received Nov. 17, 2019; revised Jan. 7, 2020; accepted Jan. 16, 2020.

The laboratory of Dr. Jean-Claude Béïque is funded by the Natural Sciences and Engineering Research Council of Canada, the Canadian Institutes of Health Research, the Heart and Stroke Foundation Canadian Partnership for Stroke Recovery, and the University of Ottawa Brain and Mind Research Institute. We thank Dr. Jean-Claude Béïque for providing a training environment that is simultaneously nurturing, exciting, and educational. We also thank Dr. Leonard Maler and Dr. Vincent Bergeron for useful discussions and comments.

*S.M. and M.L. contributed equally to this work.

The authors declare no competing financial interests.

Correspondence should be addressed to Sébastien Maillé at smail031@uottawa.ca or Michael Lynn at mlynn101@uottawa.ca.

https://doi.org/10.1523/JNEUROSCI.2740-19.2020

Copyright © 2020 the authors

cording to IIT, an AND gate possesses a value of Φ that is more than threefold higher than that of an XOR gate ($\Phi = 0.78$ vs $\Phi = 0.25$, respectively). Noel et al. (2019) did not find large enough pools of integrative and convergent neurons in vPM to generate sufficient statistical power, so this analysis was restricted to S1 neurons. The authors reasoned that if integrative neurons underlie conscious perception, then their multisensory representations should more closely track the state of consciousness than those of convergent neurons. Contrary to this, 69% of convergent neurons but only 37.1% of integrative neurons changed their multisensory response classification after propofol administration. Noel et al. (2019) additionally considered single-neuron physiological properties, including Lempel–Ziv complexity (a measure of the statistical complexity of stimulus-driven responses) and noise correlations (the amount of shared response variability between neurons). They found that both of these measures were less correlated with consciousness state in integrative neurons than in convergent neurons. Together, these findings argue against the IIT theory of consciousness.

Noel et al. (2019) also considered the following prediction of the GNW: that consciously perceived stimuli will generate a wave of activation that simultaneously spreads across multiple brain areas (neural ignition). Supporting the GNW, the authors found more coactivation of S1 and vPM during conscious awareness (beyond what would be expected from increased activation of each individual area). Based on these results, Noel et al. (2019) concluded that the data support the GNW model over IIT at the level of single neurons.

Should GNW and IIT be viewed as strictly nonoverlapping hypotheses about how neural circuits implement consciousness? Placing the findings by Noel et al. (2019) in a broader context, it is possible that the two theories are not inherently incompatible, and that GNW and IIT could provide a powerful, overlapping explanation of conscious experience at the level of anatomy and connectivity structure, respectively. Indeed, the current findings, taken at face value, simply rule out an anatomical locus (S1) and spatial scale (single neurons) at which IIT might be expressed. Proponents of IIT could point out that "integrative" neurons in this study might represent local, but not absolute maxima of integrative power; the real locus of consciousness, then, would be located in an area or at a spatiotemporal grain not recorded here. Thus, testing the IIT model and identifying the substrate of consciousness proposed by Tononi (2004) will likely require advanced techniques, including simultaneous recording and manipulation of thousands of neurons in behaving animals (Jun et al., 2017; Marshel et al., 2019). In the future, these approaches could potentially help to reconcile the GNW and IIT by defining distinct levels of abstraction for each.

One caveat of this reconciliatory perspective is that it must contend with evidence showing that primary sensory cortices can, indeed, reflect conscious processes. Kulics (1982) recorded from monkey S1 in a go/no-go task and found that a late component (~50 ms) of the evoked potential correlated with perception. Similar perception-related delayed responses have been observed in human primary auditory cortex (Wiegand and Gutschalk, 2012) and visual cortex (Maier et al., 2008). Intriguingly, anatomical data suggested that this later component likely resulted from synaptic input to cortical layers I/II, suggesting that it was generated by corticocortical loops rather than bottom-up sensory stimulation (Cauller and Kulics, 1991). These loops have been suggested to be a general feature of conscious perception (Meyer, 2011). If S1 indeed supports consciousness, and Noel et al. (2019) failed to find support for IIT in this area, one might argue that this invalidates IIT itself. However, the findings of Noel et al. (2019) are also compatible with a perspective in which primary sensory cortices simply track conscious experiences through delayed feedback projections from corticocortical loops, but cannot actually generate consciousness. This would explain why single integrative neurons are not strongly affected by anesthesia.

The work by Noel et al. (2019) also contributes to an important debate about the spatial scale over which conscious processes occur. Earlier work reported that single-unit activity in monkey somatosensory cortex did not covary with perceptual reports, suggesting that consciousness may only be present in network readouts (de Lafuente and Romo, 2005). Further evidence came from work in monkey V1 (Maier et al., 2008) showing that spiking behavior of neurons did not track perception of visual stimuli, yet BOLD (blood oxygenation level-dependent) responses and low-frequency local field potentials did. However, recent work in medial frontal cortex of humans has highlighted that the responses of individual cells can, in fact, be correlated with conscious reporting of stimuli (Reber et al., 2017; Gelbard-Sagiv et al., 2018). The study by Noel et al. (2019) helps to reconcile these findings by demonstrating that single-neuron responses are heterogeneous and dependent on information processing type: while some neurons (convergent) mostly track consciousness state, other neuron types (integrative) are seemingly not altered to the same extent. However, this also demonstrates the difficulty of inferring network properties from mixed single-unit activity, implying that perhaps a direct comparison of integrative and convergent neurons is not the ideal spatial scale for testing the IIT framework. Further work is needed to define the input–output architecture of such consciousness-tracking neurons, as well as to link the heterogeneous activity of single neurons with more abstract metrics of brain activity shown to be reflective of conscious processes. Relevant metrics might include criticality, which quantifies the extent to which a system is near the inflection point between stability and disorder (Alonso et al., 2019) and dynamic signal coordination, which quantifies the patterns of phase correlations between brain areas (Demertzi et al., 2019).

Consciousness remains one of the great unsolved mysteries in systems neuroscience. The work of Noel et al. (2019) provides an incisive direct comparison of two major theories of consciousness within the same single-unit dataset, ruling out a spatial scale and anatomical locus over which IIT formalism could potentially explain consciousness. This exciting field will, we suspect, continue to work toward a more detailed understanding of how the connectivity structure of the brain interacts with spatiotemporal activity patterns to generate human consciousness.

References

Alonso LM, Solovey G, Yanagawa T, Proekt A, Cecchi GA, Magnasco MO (2019) Single-trial classification of awareness state during anesthesia by measuring critical dynamics of global brain activity. Sci Rep 9:4927.

Cauller LJ, Kulics AT (1991) The neural basis of the behaviorally relevant N1 component of the somatosensory-evoked potential in S1 cortex of awake monkeys: evidence that backward cortical projections signal conscious touch sensation. Exp Brain Res 84:607–619.

Dehaene S, Changeux JP (2011) Experimental and theoretical approaches to conscious processing. Neuron 70:200–227.

de Lafuente V, Romo R (2005) Neuronal correlates of subjective sensory experience. Nat Neurosci 8:1698–1703.

Demertzi A, Tagliazucchi E, Dehaene S, Deco G, Barttfeld P, Raimondo F, Martial C, Fernández-Espejo D, Rohaut B, Voss HU, Schiff ND, Owen AM, Laureys S, Naccache L, Sitt JD (2019) Human consciousness is supported by dynamic complex patterns of brain signal coordination. Sci Adv 5:eaat7603.

Gelbard-Sagiv H, Mudrik L, Hill MR, Koch C, Fried I (2018) Human single neuron activity precedes emergence of conscious perception. Nat Commun 9:2057.

Ishizawa Y, Ahmed OJ, Patel SR, Gale JT, Sierra-Mercado D, Brown EN, Eskandar EN (2016) Dynamics of propofol-induced loss of consciousness across primate neocortex. J Neurosci 36:7718–7726.

Jun JJ, Steinmetz NA, Siegle JH, Denman DJ, Bauza M, Barbarits B, Lee AK, Anastassiou CA, Andrei A, Aydın Ç, Barbic M, Blanche TJ, Bonin V, Couto J, Dutta B, Gratiy SL, Gutnisky DA, Häusser M, Karsh B, Ledochowitsch P, et al. (2017) Fully integrated silicon probes for high-density recording of neural activity. Nature 551:232–236.

Kulics AT (1982) Cortical neural evoked correlates of somatosensory stimulus detection in the rhesus monkey. Electroencephalogr Clin Neurophysiol 53:78–93.

Lemon RN, Edgley SA (2010) Life without a cerebellum. Brain 133:652–654.

Maier A, Wilke M, Aura C, Zhu C, Ye FQ, Leopold DA (2008) Divergence of fMRI and neural signals in V1 during perceptual suppression in the awake monkey. Nat Neurosci 11:1193–1200.

Marshel JH, Kim YS, Machado TA, Quirin S, Benson B, Kadmon J, Raja C, Chibukhchyan A, Ramakrishnan C, Inoue M, Shane JC, McKnight DJ, Yoshizawa S, Kato HE, Ganguli S, Deisseroth K (2019) Cortical layer-specific critical dynamics triggering perception. Science 365:eaaw5202.

Meyer K (2011) Primary sensory cortices, top-down projections and conscious experience. Prog Neurobiol 94:408–417.

Noel JP, Ishizawa Y, Patel SR, Eskandar EN, Wallace MT (2019) Leveraging nonhuman primate multisensory neurons and circuits in assessing consciousness theory. J Neurosci 39:7485–7500.

Reber TP, Faber J, Niediek J, Boström J, Elger CE, Mormann F (2017) Single-neuron correlates of conscious perception in the human medial temporal lobe. Curr Biol 27:2991–2998.e2.

Tononi G (2004) An information integration theory of consciousness. BMC Neurosci 5:42.

Tononi G, Boly M, Massimini M, Koch C (2016) Integrated information theory: from consciousness to its physical substrate. Nat Rev Neurosci 17:450–461.

Wiegand K, Gutschalk A (2012) Correlates of perceptual awareness in human primary auditory cortex revealed by an informational masking experiment. Neuroimage 61:62–69.

Yu F, Jiang QJ, Sun XY, Zhang RW (2015) A new case of complete primary cerebellar agenesis: clinical and imaging findings in a living patient. Brain 138:e353.