

## Letter

# Scientific communication and the semantics of sentience

Brett J. Kagan,<sup>1,11,\*</sup> Adeel Razi,<sup>4,6,7,8</sup> Anjali Bhat,<sup>4</sup> Andy C. Kitchen,<sup>1</sup> Nhi T. Tran,<sup>2</sup> Forough Habibollahi,<sup>9</sup> Moein Khajehnejad,<sup>10</sup> Bradyn J. Parker,<sup>3</sup> Ben Rollo,<sup>5</sup> and Karl J. Friston<sup>4</sup>

<sup>1</sup>Cortical Labs, Melbourne, VIC, Australia

<sup>2</sup>The Ritchie Centre, Hudson Institute of Medical Research, Clayton, VIC, Australia

<sup>3</sup>Monash University, Melbourne, VIC, Australia

<sup>4</sup>Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, WC1N 3BG, UK

<sup>5</sup>Department of Neuroscience, Central Clinical School, Monash University, Melbourne, VIC, Australia

<sup>6</sup>Turner Institute for Brain and Mental Health, Monash University, Clayton, VIC, Australia

<sup>7</sup>Monash Biomedical Imaging, Monash University, Clayton, VIC, Australia

<sup>8</sup>CIFAR Azrieli Global Scholars Program, CIFAR, Toronto, ON, Canada

<sup>9</sup>Department of Biomedical Engineering, The University of Melbourne, Parkville, VIC, Australia

<sup>10</sup>Department of Data Science and AI, Faculty of IT, Monash University, Melbourne, VIC, Australia

<sup>11</sup>Lead contact

\*Correspondence: [brett@corticallabs.com](mailto:brett@corticallabs.com)

<https://doi.org/10.1016/j.neuron.2023.02.008>

The use of language to describe specific phenomena has always been, and will likely remain, a contentious aspect of scientific discourse. Effective scientific communication must be considered in the context of the field in which the given signifiers are used. While the response by Balci et al. may be described as polemic and contains reasoning subject to equivocation and other fallacies, we do appreciate the core concerns. Here, we address these concerns and suggest some constructive pathways that might help improve scientific communication in the future.

Firstly, the assumption that we acted to internationally “oversell” and create “media buzz” or “controversy” should be addressed. Contrary to this futile fallacy, our intention in using “sentience” and related terminology was to be deflationary and principled—in line with the recent literature in theoretical biology in general and in the free energy principle (FEP) in particular. The FEP is a first principles account of biotic self-organization, namely, active inference or sentient behavior.<sup>1</sup> In this setting, many terms acquire a straightforward and technical meaning; for example, “surprise” refers to self-information, “uncertainty” to information entropy, and Bayesian “beliefs” to posterior probability distributions<sup>1</sup>—in the sense of Bayesian belief updating and propagation. This reflects a recent trend in treatments of self-organization that tries to find the common ground in

physics, biology, and psychology, e.g., basal cognition.<sup>2</sup> The intent is to concretize concepts in testable and falsifiable ways<sup>3</sup>—and in a way that underwrites current collaborations testing theories of consciousness, e.g., <https://www.templetonworldcharity.org/projects-database/0646>.

The core results that were presented<sup>4</sup> have been publicly available via a preprint on bioRxiv since 2021, allowing widespread, careful consideration, collegiate communication, and outreach to the wider scientific community. The overwhelmingly helpful and supportive comments received led to significant refinement, including a clear definition—in the peer-reviewed version—of “sentience.” This definition reflects current philosophical accounts of sentience in predictive processing (e.g., <https://www.penguinrandomhouse.com/books/608016/the-experience-machine-by-andy-clark/>). Furthermore, the suggestion by Balci et al. that our use of terms is “unsupported” is inaccurate. Journal requirements for these letters (limited to seven references) precludes a detailed inclusion of the supporting literature. Yet, as an example, recent rigorous endeavors to study the biological basis of sentience have found our formal definitions useful.<sup>3</sup> More generally, at least 71 distinct definitions of “intelligence” have been previously identified, and our usage accords with many of these definitions.<sup>5</sup> However, leaving aside the above-mentioned concerns,

we now focus on common ground and constructive pathways forward.

Foremost, we agree with the concluding sentiment by Balci et al. that the questions and challenges around definitions of intelligent behavior and sentience are important. The widespread positive (sentiment analysis shows only <9% to be negative, which includes outlandish concerns such as the creation of “zombies”) response to our work highlights the importance of establishing a shared nomenclature to communicate results clearly. In the short term, we propose that future work should include (and we would encourage journals to allow) glossaries. Glossaries provide clarity for the reader, in terms of understanding what a particular term is meant to convey. This approach has been adopted in an upcoming work on which some of our authors collaborated.<sup>6</sup> In the longer term, it would be beneficial to establish a generally accepted nomenclature for standard definitions. We have made open invitations to the scientific community for collaboration and take this opportunity to again welcome collaborative interest. Especially as, poignantly, the current exchange highlights the importance of good-faith discussions in formulating nomenclature standards.

Secondly, we agree with Balci et al. that accurate scientific communication is important. For our work, we took care not to oversell the research and minimize hype often seen in related fields. We made no claims about treatments or cures

for disease. We circulated the media release to experienced science reporters via EurekAlert and carefully briefed experienced science reporters to promote responsible reporting. A primary motive for creating *DishBrain* is to study how genetic alterations and pharmacological agents influence the real-time behavior of neurons—to elucidate underlying mechanisms of interest. Any discussions of more complex applications were carefully positioned as future work, requiring further research and development. Likewise, we do acknowledge that citations identified by Balci et al. were not present in our introduction. As with most research articles, due to space constraints our review of the literature was not intended to be exhaustive and required highlighting only a subset of previous work in this field. However, we do acknowledge the need for future work to focus on consolidating the cross-disciplinary work in this field.

Finally, we also agree with Balci et al. around the importance of making sure that ethical debates are not fueled by misunderstandings. To that extent, we have engaged with independent ethicists, discussing terms such as “sentience” in this context<sup>7</sup> and aiming to further explore these issues more in the future. While it is reasonable to note that improvements to this technology could, for some applications, offer an

alternative to behavioral research using animals, at no point do we imply animal research can be completely replaced. Yet this does not mean the possibility of advancements should be discounted or that discussions about future applications of scientific work are inherently “overselling.” New technologies, such as synthetic biological intelligence, have the potential to offer significant gains to both society and science if we can work collaboratively to realize the potential benefits. Explaining this potential is an important part of scientific communication to frame where future work may lead. Ultimately, as much as signifiers in scientific communication must be understood in context, the scientific outcomes must also be considered in their own context—in this case, as *in vitro* work. Fundamental exhibitions of any phenomena in a dish may be hard to define, especially without consistent definitions within and across fields, as we see here. Multidisciplinary work seeking to advance research is seldom likely to be perfect initially due to difficulties integrating across fields. Nomenclature may differ, and approaches diverge. Yet, if these differences are recognized not only as an opportunity to criticize but to improve and innovate, our scientific community and scientific progress can hopefully continue to benefit.

## REFERENCES

1. Friston, K.J., Wiese, W., and Hobson, J.A. (2020). Sentience and the Origins of Consciousness: From Cartesian Duality to Markovian Monism. *Entropy* 22, 516. <https://doi.org/10.3390/e22050516>.
2. Fields, C., Glazebrook, J.F., and Levin, M. (2021). Minimal physicalism as a scale-free substrate for cognition and consciousness. *Neurosci. Conscious.* 2021, niab013. <https://doi.org/10.1093/nc/niab013>.
3. Pereira, A., Garcia, J.W., and Muotri, A. (2023). Neural Stimulation of Brain Organoids with Dynamic Patterns: A Sentiomics Approach Directed to Regenerative Neuromedicine. *NeuroSci* 4, 31–42. <https://doi.org/10.3390/neurosci4010004>.
4. Kagan, B.J., Kitchen, A.C., Tran, N.T., Habibollahi, F., Khajehnejad, M., Parker, B.J., Bhat, A., Rollo, B., Razi, A., and Friston, K.J. (2022). *In vitro* neurons learn and exhibit sentience when embodied in a simulated game-world. *Neuron* 110, 3952–3969.e8. <https://doi.org/10.1016/j.neuron.2022.09.001>.
5. Legg, S., and Hutter, M. (2007). A collection of definitions of intelligence. *Frontiers in Artificial Intelligence and applications* 157, 17.
6. Smirnova, L., Caffo, B.S., Gracias, D.H., Huang, Q., Morales Pantoja, I.E., Tang, B., Zack, D.J., Berlinicke, C.A., Boyd, L., Harris, T., et al. (2023). Organoid intelligence (OI): the new Frontier in biocomputing and intelligence-in-a-dish. *Front. Sci.* 1. <https://doi.org/10.3389/fsci.2023.1017235>.
7. Kagan, B.J., Duc, D., Stevens, I., and Gilbert, F. (2022). Neurons Embodied in a Virtual World: Evidence for Organoid Ethics? *AJOB Neurosci.* 73, 114–117. <https://doi.org/10.1080/21507740.2022.2048731>.