Royal Netherlands Academy of Arts and Sciences (KNAW)
KONINKLIJKE NEDERLANDSE AKADEMIE VAN WETENSCHAPPEN

A response to claims of emergent intelligence and sentience in a dish

Balci, Fuat; Ben Hamed, Suliann; Boraud, Thomas; Bouret, Sébastien; Brochier, Thomas; Brun, Cédric; Cohen, Jeremiah Y; Coutureau, Etienne; Deffains, Marc; Doyère, Valérie; Gregoriou, Georgia G; Heimel, J Alexander; Kilavik, Bjørg Elisabeth; Lee, Daeyeol; Leuthardt, Eric C; Mainen, Zachary F; Mathis, Mackenzie; Monosov, Ilya E; Naudé, Jérémie; Orsborn, Amy L; Padoa-Schioppa, Camillo; Procyk, Emmanuel; Sabatini, Bernardo; Sallet, Jérôme; Sandi, Carmen; Schall, Jeffrey D; Soltani, Alireza; Svoboda, Karel; Wilson, Charles R E; Zimmermann, Jan

published in
Neuron
2023

DOI (link to publisher)
10.1016/j.neuron.2023.02.009

document version
Publisher's PDF, also known as Version of record

Link to publication in KNAW Research Portal

citation for published version (APA)
Balci, F., Ben Hamed, S., Boraud, T., Bouret, S., Brochier, T., Brun, C., Cohen, J. Y., Coutureau, E., Deffains, M., Doyère, V., Gregoriou, G. G., Heimel, J. A., Kilavik, B. E., Lee, D., Leuthardt, E. C., Mainen, Z. F., Mathis, M., Monosov, I. E., Naudé, J., ... Zimmermann, J. (2023). A response to claims of emergent intelligence and sentience in a dish. *Neuron*, *111*(5), 604-605. https://doi.org/10.1016/j.neuron.2023.02.009

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the KNAW public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain.
• You may freely distribute the URL identifying the publication in the KNAW public portal.

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:
pure@knaw.nl

Download date: 07. Aug. 2024

CellPress

Neuron

Letter

A response to claims of emergent intelligence and sentience in a dish

Fuat Balci,[1] Suliann Ben Hamed,[2] Thomas Boraud,[3,4] Sébastien Bouret,[5] Thomas Brochier,[6] Cédric Brun,[3,7] Jeremiah Y. Cohen,[8] Etienne Coutureau,[9] Marc Deffains,[3] Valérie Doyère,[10] Georgia G. Gregoriou,[11] J. Alexander Heimel,[12] Bjørg Elisabeth Kilavik,[6] Daeyeol Lee,[13] Eric C. Leuthardt,[14] Zachary F. Mainen,[15] Mackenzie Mathis,[16] Ilya E. Monosov,[17,*] Jérémie Naudé,[18] Amy L. Orsborn,[19] Camillo Padoa-Schioppa,[17] Emmanuel Procyk,[20,*] Bernardo Sabatini,[21] Jérôme Sallet,[20] Carmen Sandi,[16] Jeffrey D. Schall,[22] Alireza Soltani,[23] Karel Svoboda,[8] Charles R.E. Wilson,[20] and Jan Zimmermann[24]

[1]University of Manitoba, Department of Biological Sciences, 50 Sifton Rd, Winnipeg, MB R3T 2M5, Canada
[2]Institut des Sciences Cognitives Marc Jeannerod, UMR5229 CNRS Université de Lyon, 67 Boulevard Pinel, 69675 Bron Cedex, France
[3]University Bordeaux, CNRS, IMN, UMR 5293, 33000 Bordeaux, France
[4]CHU de Bordeaux, IMN Clinique, 33000 Bordeaux, France
[5]Institut du Cerveau (ICM), INSERM UMRS 1127, CNRS UMR 7225, Pitié-Salpêtrière Hospital, Paris, France
[6]Institut de Neurosciences de la Timone (INT), UMR 7289, CNRS, Aix-Marseille Université, Marseille 13005, France
[7]University Bordeaux Montaigne, Philosophy Department, 33607 Pessac, France
[8]Allen Institute for Neural Dynamics, Seattle, WA 98109, USA
[9]University Bordeaux, CNRS, INCIA, UMR 5287, 33000 Bordeaux, France
[10]Université Paris-Saclay, CNRS, Institut des Neurosciences Paris-Saclay, UMR9197, 91400 Saclay, France
[11]University of Crete, Faculty of Medicine and Foundation for Research and Technology Hellas, Institute of Applied and Computational Mathematics, Heraklion 70013, Greece
[12]Netherlands Institute for Neuroscience, Amsterdam, the Netherlands
[13]Zanvyl Krieger Mind/Brain Institute, Departments of Psychological and Brain Sciences and Neuroscience, Johns Hopkins University, Baltimore, MD, USA
[14]Department of Neurosurgery, Washington University School of Medicine, St. Louis, MO, USA
[15]Champalimaud Foundation, Lisbon, Portugal
[16]École Polytechnique Fédérale de Lausanne (EPFL), Brain Mind Institute, Lausanne, Switzerland
[17]Department of Neuroscience, Washington University School of Medicine, St. Louis, MO, USA
[18]IGF, University of Montpellier, CNRS, INSERM, Montpellier, France
[19]Department of Electrical and Computer Engineering, University of Washington, Seattle, WA, USA
[20]University Lyon, Université Lyon 1, Inserm, Stem Cell and Brain Research Institute U1208, Bron, France
[21]HHMI, Department of Neurobiology, Harvard Medical School, Boston, MA, USA
[22]Visual Neurophysiology Centre, York University, Toronto, ON, Canada
[23]Department of Psychological and Brain Sciences, Dartmouth University, Hanover, NH, USA
[24]University of Minnesota, Department of Neuroscience, Center for Magnetic Resonance Research, Minneapolis, MN, USA
*Correspondence: ilya.monosov@gmail.com (I.E.M.), emmanuel.procyk@inserm.fr (E.P.)
https://doi.org/10.1016/j.neuron.2023.02.009

The article "*In vitro* neurons learn and exhibit sentience when embodied in a simulated game-world" by Kagan et al.[1] triggered a wave of positive mainstream and scientific media coverage as well as a widespread negative reaction from the scientific community. Here, we discuss why this negative reaction is legitimate and must be taken seriously. We raise concerns about the key claim of the article: that it demonstrates that "a single layer of *in vitro* cortical neurons can self-organize activity to display intelligent and sentient behavior when embodied in a simulated game-world."

Our concerns go beyond the appropriateness of the methodology, quantification, and controls used in the study and the lack of details with which they were presented. These concerns include the unsupported use of terms and concepts that misrepresent the findings of this study, the lack of acknowledgment of previous literature, and the ensuing overselling of translational and societal relevance.

The first concern regards the unsupported use of terms and concepts to describe properties of biological and/or artificial neural networks, such as "sentience," "goal-directed behavior," "embodiment," tackling "uncertainty," and "intelligence." Assessing whether animals, and their neural networks, display these properties is indeed an important topic in neurobiology and computer science, as well as in philosophy, psychology, and ethology. This effort has led to a progressive refinement of these concepts, and the development of benchmarking for such cognitive capacities has recently become a crucial aim of machine learning and artificial intelligence research. Considering this large-scale effort, we believe Kagan et al. made strong claims for the application of these terms to neural networks with relatively weak evidence.

By associating elements of cognition with the properties of cultured neurons, Kagan et al. inevitably created a media buzz. This paper also attracted a lot of attention in the scientific community in part because of concerns about "concept hijacking" or at least their misleading usage. Although there are indeed legitimate discussions in the field about how to benchmark and test the abilities of agents and networks to display cognitive or goal-directed behaviors (and about how to precisely define such behaviors), the current report does not evaluate the outcome of their experimental observations on those

grounds and, in addition, makes claims well beyond the acquired data and effect sizes.

For example, attributing intelligence to a network that displays short-term plasticity is not supported by relevant scientific fields such as machine learning, neurobiology, and psychology. Similar arguments apply to the yet more provocative use of the term sentience. This application to neurons *in vitro* is in our view even more inappropriate and is not justified by the data presented in the paper. The term sentience is notoriously hard to define but refers to a process that encompasses feeling, sensing, and subjective evaluation.[2] The application of intelligence and sentience to neurons-in-a-dish in this paper is not based on any established or robust consensus on the definitions of these very important terms.[3] Instead, it is based on the authors' own recent theoretical propositions,[4] which are general enough to allow the term to be applied to nearly any interactive computational system of even modest complexity.

Beyond provoking a controversy, it is unclear how the use of terms such as sentience and intelligence adds to the understanding of neural network properties in this paper. Because it is currently challenging for mechanistic or reductionist neurobiological studies to link these concepts to biological phenomena, we suggest that the terms ought to be used with more caution. Moreover, the concept of sentience has a key role in the philosophical and sociological issues surrounding animal welfare and for that reason should not be used loosely or in unconventional manners in the context of this or any other scientific study.

To be clear, we are not arguing that research in isolated neural networks is problematic. Unquestionably, these approaches can provide crucial knowledge of neural network dynamics, plasticity, and computational and organizational principles and processing capacities. In fact, beyond the unnecessary or unfounded use of terminology, further concerning are limitations of results and failures of scholarship. Strong conclusions are compromised by weak results, some of which fail to adequately match control and experimental conditions. Also, Kagan et al. do not acknowledge previous use of biological neural networks embedded in closed-loop systems that has helped, for example, to assess the potential application of plasticity to drive external artifacts, e.g., robots.[5,6]

We conclude our opinion with a discussion of why this paper and the media coverage of it illustrate the importance of scientific communication to the general population. Media tend to directly republish information included in abstracts and significance statements, and interviews of scientists by media tend to amplify these statements. Overselling scientific results directly impacts the evaluation of scientific reliability and credibility.[7] In the present specific case, claiming that a cell culture embedded in a closed-loop system demonstrates sentience and intelligence might impact the public perception of what in nature is sentient and intelligent and could trigger ethical debates fueled by misunderstanding. It puts an unnecessary risk on the whole community of systems neuroscience that tries to understand higher brain functions and dysfunctions by fueling an argument, albeit invalid, to extreme animal rights movements that lobby daily to stop animal research, while also creating potential future financial benefits for the possible usage of the methods in this paper.

Studies related to nervous systems and their computational abilities represent a huge area of research for the advancement of our knowledge of what we are and what we are capable of and accordingly pose several scientific, ethical, and societal challenges. Therefore, the questions and challenges we raise regarding definitions of intelligent behavior and sentience in neuroscience in general, and in the work by Kagan et al. specifically, are of fundamental importance to fulfill the high expectations that neuroscience has created for understanding brain functions, curing brain diseases, and conducting responsible research in developing machines capable of performing complex behavior.

DECLARATION OF INTERESTS

The writers and key corresponding authors of this document, I.E.M. and E.P., do not have any patents or interests related to our work to declare. The rest of the authors are "signers" and gave comments to improve the manuscript, therefore they were not polled on this issue.

REFERENCES

1. 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.

2. Broom, D.M. (2016). Considering animals' feelings: Précis of Sentience and animal welfare (Broom 2014). Animal Sentience *5*. https://doi.org/10.51291/2377-7478.1015.

3. Lee, D. (2020). Birth of Intelligence: From RNA to Artificial Intelligence, First edition (Oxford University Press). https://doi.org/10.1093/oso/9780190908324.001.0001.

4. 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.

5. Tessadori, J., Bisio, M., Martinoia, S., and Chiappalone, M. (2012). Modular neuronal assemblies embodied in a closed-loop environment: toward future integration of brains and machines. Front. Neural Circuits *6*, 99. https://doi.org/10.3389/fncir.2012.00099.

6. Reger, B.D., Fleming, K.M., Sanguineti, V., Alford, S., and Mussa-Ivaldi, F.A. (2000). Connecting brains to robots: an artificial body for studying the computational properties of neural tissues. Artif. Life *6*, 307–324. https://doi.org/10.1162/106454600300103656.

7. West, J.D., and Bergstrom, C.T. (2021). Misinformation in and about science. Proc. Natl. Acad. Sci. USA *118*, e1912444117. https://doi.org/10.1073/pnas.1912444117.