Embodied Neural Systems Can Enable Iterative Investigations of Morally Relevant States

Brett J. Kagan[1, 7*], Alon Loeffler[1], J. Lomax Boyd (ORCID: 0000-0001-9878-3493)[2], Julian Savulescu[4, 5, 6]
1 Cortical Labs, Melbourne, VIC, Australia
2 Berman Institute of Bioethics, Johns Hopkins University, Baltimore, MD, United States
4 Murdoch Children's Research Institute, Melbourne, VIC, Australia
5 Melbourne Law School, University of Melbourne, Melbourne, VIC, Australia
6 Centre for Biomedical Ethics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore
7 Department of Biochemistry and Pharmacology, University of Melbourne, Parkville, VIC, Australia.
**Corresponding Author: brett@corticallabs.com*

The development of new technologies is best accompanied by—ideally, preceded by—serious ethical reflection. Perhaps this is because only after integrating interdisciplinary expertise is it possible to fully capture the breadth of issues regarding how any new technology may intersect with society (Savulescu, 2015). The development of intelligent information processing systems that integrate neural tissue with synthetic computing systems, termed Synthetic Biological Intelligence (SBI) systems, is a technology that would benefit from integration of both scientific and ethical expertise. The uncertainties of developing systems that may possess potentially morally relevant states (e.g., phenomenological consciousness—the unique subjective quality associated with an experience) prompts discussion about how to identify when such states arise and, if identified, what considerations are due. Here we propose that understanding the moral status of complicated dynamic phenomena that make up neural systems (whether animal or human) is best done by building up simple tests to identify clear metrics of interest across systems that possess multiple levels of complexity.

Research into how neural systems produce intelligent phenomena is progressing rapidly. Findings that *in vitro* neural systems may be able to engage in logical and complex behaviours has further spurred development (reviewed in (Kagan et al., 2023)). An example of such intelligent phenomena was demonstrated using *in vitro* cortical cells that displayed basic learning within a real-time, closed-loop system within a short time (Kagan et al., 2022). Many researchers are already seeking to further expand the complexities of these SBI systems, for example, exploring so-called in 3-dimensional brain organoids, (Hartung et al., 2023). As this new technology progresses rapidly, it is crucial for ethical standards to keep pace.

Such ethical considerations will require an iterative process of development and testing of neural systems, not only at a cellular level, but also within entire biological systems (such as in a human). As a key example of the morally relevant states identified in the literature, we focus here on phenomenological consciousness. However, the rationale and arguments are applicable across any state of interest.

1. Neuroscience informing ethics

The rapid growth of interest in SBI, coupled with the array of potential applications (Kagan et al., 2023), limited agreement on nomenclature (Pereira et al., 2023), and tricky philosophical questions regarding what systems even qualify for moral status (Boyd and Lipshitz, 2023), cements the need for interdisciplinary consideration of how to ethically investigate SBI phenomena. Many of the potential concerns overlap with other technologies such as stem-cell therapy (e.g., donor rights, privacy, etc.), or with other silicon-based technologies that may lead to diverse intelligent systems (e.g., the possibility of automation limiting jobs, equity of access, etc.). Yet the inclusion of biological—typically human—tissue has raised the specific query of whether SBI-type systems could possess morally relevant states (Goddard et al., 2023). While other phenomena exist that may allow a system to be considered to possess morally relevant states, such as intelligence or capacity for suffering, many consider phenomenological consciousness to be a crucial factor (Goddard et al., 2023). While no evidence that simple *in vitro* systems can possess morally relevant traits exists, there are already calls for objective metrics to identify morally relevant traits (Kagan et al., 2023; Pereira et al., 2023). Yet our understanding of the biological underpinnings of these traits lags the recognition of their potential ethical value.

The challenge of identifying metrics to indicate where morally relevant states arise is that the emergence of any complex traits from a neural system likely stems from a part-whole relationship within the wider dynamic system (Mediano et al., 2022). To measure any given neural correlate in a complex organism, one would need to separate the activity associated with consciousness from concurrent activity associated with that correlate—a difficult, nigh impossible task. At minimum, processes are necessary to communicate an experience to others. The use of language regions, or motor regions for non-linguistic reports, will result in neural activity. At best, this increases noise, and at worst, it confounds experimental interpretations. Other processes including attention and memory further complicate these interpretive efforts. Perhaps for the above reasons, attempts to isolate neural correlates of consciousness, even when supportive of specific theories, have yet to achieve a consensus about what features or processes may give rise to phenomenological consciousness. Thus, there is a scientific need to disassemble neural systems into their constitute parts so that 'intelligent' systems can be built-up (and understood) from principal components.

SBI systems are far simpler than human or animal subjects for modelling biological processes that contribute to morally relevant traits. Even if the system under investigation is so simple that it would be considered a non-conscious system by most, it holds merit in establishing what metrics can arise without complex neural structures. For example, neural criticality—activity patterns balanced between ordered and disordered states—has previously been proposed to support consciousness and to be a useful index of consciousness in clinical settings (Walter and Hinterberger, 2022). Yet, critical dynamics were also found to arise in monolayers of neurons *in vitro*, especially when placed in a structured information landscape through real-time, closed-loop stimulation and recordings to simulate a game (Habibollahi et al., 2023). This highlights that, while close-to-critical dynamics arise when organisms are taking in structured information, such dynamics do not require higher order cognition or complex neural structures. Therefore, it is reasonable to argue that neural criticality as a metric is unsuitable as a *standalone* marker for anything as complex as phenomenological consciousness. In this example, criticality is valid in a clinical setting and links to the phenomena of interest, yet adding additional simplified modes of investigation provide a different and more nuanced (albeit deflationary) understanding.

2. Ethics informing neuroscience

For appropriate ethical treatment of SBI systems to become possible, gaining a clearer understanding of where morally relevant states arise within them is necessary. For this understanding to develop, coordination of empirical investigation across systems of varied complexity from the whole organism to the single cell is required. Technological advances, such as SBI and OI, offer an assortment of tools for such exploration at lower structural complexities.

By ensuring that theories that propose to explain or predict phenomena of interest (whether they are signified as “consciousness”, “intelligence”, “sentience”, etc.) have testable implications that can be applied (to some extent) to different levels of testing (e.g. to human subjects, rodents, *in vitro*), it may become possible to identify not only where the metrics arise, but also where they do not. Previously, it was not feasible to test most meaningful theories *in vitro* due to technological limitations. However, the new technologies that enable the embodiment of simple cultures in structured information landscapes allow far more nuanced investigation at this level for morally relevant traits. For example, should a metric proposed as a marker of a phenomenon be found in humans, but not in animal or *in vitro*, despite rigorous investigation, this may indicate that certain structures unique to humans are required to give rise to the phenomena of interest. Such a conclusion cannot be reached by simply finding and identifying a phenomenon in humans alone without a comparison system of varied complexity. Conversely, if adding complexity to the biological material, or altering the structure of stimulation, can give rise to a phenomenon observed in humans, but only under certain conditions within a cell culture, this would also start to identify the requirements for the phenomenon of interest; a process which could be termed ‘experimental neuroethics’.

3. Interdisciplinary collaboration

An iterative process to uncover the scientific and ethical complexity of biological systems will require a range of skilled professionals from multiple disciplines who may rarely interact. To gain a better understanding of where markers for morally relevant states arise, in which context, and what ethical ramifications may be associated, open communication and collaboration among representative stakeholders is required. This would foster a comprehensive approach to understanding morally relevant states from the cellular level all the way up to the intact brain, as well as any ethical considerations that may arise anywhere in between. It would also prompt necessary discussion around a universally agreed-upon nomenclature surrounding research into relevant terms as previously proposed (Kagan et al., 2023).

Therefore, an iterative process of development and testing offers three key benefits. Firstly, it provides a new tool to investigate metrics that may predict or correlate with complex, morally relevant states. In this way, neuroscience can inform ethics. Secondly, it will aid in identifying when increasingly complicated *in vitro* cultures may start to show hallmarks of potentially morally relevant states, which would necessitate more rigorous ethical consideration, thereby allowing ethics to inform neuroscience. Thirdly, this approach may foster a greater degree of ongoing active collaboration between different disciplines, such as scientists and ethicists, which contributes to the kind of interdisciplinary research advocated for by science agencies. Taking these three advantages together, this iterative process of integrating established theories with new technologies has the potential to accelerate our understanding of the mechanisms that allow the wonderfully complex nature facilitated by the brain to arise.

Acknowledgements

The authors thank and acknowledge Professor Anil Seth and Dr Nhi Thao Tran for their guidance and contributions to the manuscript. Professor Julian Savulescu, through his involvement with the Murdoch Children's Research Institute, received funding through from the Victorian State Government through the Operational Infrastructure Support (OIS) Program. This research is supported by the Singapore Ministry of Health's National Medical Research Council under its Enablers and Infrastructure Support for Clinical Trials-related Activities Funding Initiative (NMRC Project No. MOH-000951-00). Professor Lomax Boyd is supported by the Kavli Foundation and SURPASS award from the Johns Hopkins Whiting School of Engineering and Applied Physics Laboratory.

Potential Conflict of Interests

B.J.K. and A.L. are employees of Cortical Labs which does work related to the subject matter in the manuscript and hold an interest in patents related to the subject matter in the manuscript. B.J.K. holds shares in Cortical Labs.

References

Boyd JL, Lipshitz N (2023) Dimensions of Consciousness and the Moral Status of Brain Organoids. Neuroethics 17:5.

Goddard E, Tomaskovic-Crook E, Crook JM, Dodds S (2023) Human Brain Organoids and Consciousness: Moral Claims and Epistemic Uncertainty. Organoids 2:50–65.

Habibollahi F, Kagan BJ, Burkitt AN, French C (2023) Critical dynamics arise during structured information presentation within embodied in vitro neuronal networks. Nat Commun 14:5287.

Hartung T et al. (2023) The Baltimore declaration toward the exploration of organoid intelligence. Front Sci 1:1068159.

Kagan BJ, Gyngell C, Lysaght T, Cole VM, Sawai T, Savulescu J (2023) The technology, opportunities, and challenges of Synthetic Biological Intelligence. Biotechnology Advances 68:108233.

Kagan BJ, Kitchen AC, Tran NT, Habibollahi F, Khajehnejad M, Parker BJ, Bhat A, Rollo B, Razi A, Friston KJ (2022) In vitro neurons learn and exhibit sentience when embodied in a simulated game-world. Neuron:S0896627322008066.

Mediano PAM, Rosas FE, Luppi AI, Jensen HJ, Seth AK, Barrett AB, Carhart-Harris RL, Bor D (2022) Greater than the parts: a review of the information decomposition approach to causal emergence. Phil Trans R Soc A 380:20210246.

Pereira A, Garcia JW, Muotri A (2023) Neural Stimulation of Brain Organoids with Dynamic Patterns: A Sentiomics Approach Directed to Regenerative Neuromedicine. NeuroSci 4:31–42.

Savulescu J (2015) Bioethics: why philosophy is essential for progress. J Med Ethics 41:28–33.

Walter N, Hinterberger T (2022) Self-organized criticality as a framework for consciousness: A review study. Front Psychol 13:911620.