



Levin, Michael (2022) [Generalizing frameworks for sentience beyond natural species](#). *Animal Sentience* 32(15)

DOI: 10.51291/2377-7478.1733

Date of submission: 2022-06-26

Date of acceptance: 2022-06-27



This article has appeared in the journal *Animal Sentience*, a peer-reviewed journal on animal cognition and feeling. It has been made open access, free for all, by WellBeing International and deposited in the WBI Studies Repository. For more information, please contact wbisr-info@wellbeingintl.org.



Generalizing frameworks for sentience beyond natural species

Commentary on [Crump et al.](#) on *Decapod Sentience*

Michael Levin

Allen Discovery Center, and Department of Biology, Tufts University

Abstract: Crump et al. (2022) offer a well-argued example of an essential development: a rigorous framework for assessing sentience from the perspective of moral concern over an agent's welfare. Current and forthcoming developments in bioengineering, synthetic morphology, artificial intelligence, biorobotics, and exobiology necessitate an expansion and generalization of this effort. Verbal reports (the Turing Test) and homology to human brains are utterly inadequate criteria for assessing the status of novel, unconventional agents that offer no familiar touchstone of phylogeny or anatomy. We must develop principled approaches to evaluating the sentience of (and thus, our responsibility to) beings of unfamiliar provenance and composition.

[Michael Levin](#), Director of the Allen Discovery Center at Tufts University, does research at the intersection of developmental biology, computer science, and cognitive science.

[Website](#)



1. Context and scope: evaluating sentience. Crump et al. (2022) provide a welcome contribution to a pressing problem, which stands at the heart of two difficult issues occupying mankind for millennia: the problem of other minds, and the challenge of establishing instructive norms based on scientific data (“deriving *ought* from *is*”). What degree of concern and care should we exhibit toward the many diverse agents around us, and what criteria do we use to identify sentience, capacity for suffering, and other properties that have moral implications?

Crump et al.’s contribution is practical, quantitative, and grounded in empirical data. It does not address the Hard Problem – why any of the functional/structural criteria they lay out should be associated with feeling (Chalmers, 1996; Nagel, 1986), but this problem is likely unsolvable in 3rd person and should not stand as a barrier to progress. As they correctly point out, demanding certainty in this case is inappropriate, and we should err on the side of reducing false negatives with respect to sentience criteria for ethical concern.

The essential contribution of Crump et al. is to lay out an example of a set of rigorous, flexible criteria for making judgements in specific cases. Their framework is ideally adapted to a wide range of natural biological species. Moreover, they explicitly highlight the fact that the criteria for concern should not require 2nd order abilities of creatures to reflect on their feelings (an especially advanced region of the cognitive continuum). This is an important link to the field of “minimal cognition”, which studies the evolutionary origins (primitive versions) of familiar, complex cognitive capacities in models such as bacteria, slime moulds, somatic tissues, plants, etc. (Levin et al., 2021; Lyon, 2006; Lyon, 2020; Lyon et al., 2021).

Here, I briefly discuss this kind of effort from a wider perspective, beyond the ethology of natural kinds in the biosphere.

2. Endless “Forms Most Beautiful” 2.0: the space of possible beings. One of Darwin’s enormous impacts was to show how the familiar distinction of humans vs. mere animals was but a poor coarse-graining (binarization) of a *continuum*. This axis links the simplest chemical replicators to humans in a gradual, continuous manner. Similarly, developmental biology reveals how each of us, uncontroversially sentient beings, morphed slowly and gradually from the “just physics” of a quiescent oocyte cell to a human-level mind (Levin, 2019). These 1-dimensional phylogenetic and ontogenetic life histories, however, are just the beginning of a revolution that was perhaps unimaginable in Darwin’s day (Levin, 2020).

We stand at the entry point to a great transition, in which all of Darwin’s familiar “Forms Most Beautiful” (Darwin, 1859) are just a tiny region of an astronomically large space of possible beings. Already, the merger of living tissue and smart materials are giving rise to a variety of cyborgs (Orive et al., 2020; Pio-Lopez, 2021) - humans with novel sensors, prosthetic limbs and device control that implements Clark & Chalmers’s (1998) “Extended Mind” Hypothesis, and smart implants that modulate cognitive function. Chimeric technologies, biorobotics, and evolutionary strategies used in robotics are blurring the line between evolved and designed agents – organisms and machines (Bongard and Levin, 2021). Closed-loop platforms integrating both biological and machine learning components, and “hybrots” (cultured brains with robotic bodies), are revealing the proto-cognitive capacities of cells and tissues in novel configurations (Bakkum et al., 2007; DeMarse and Dockendorf, 2005; Kagan et al., 2021; Potter et al., 2005). Bioengineering and synthetic morphology are giving rise to coherent, autonomous “beings” that have anatomies, cellular composition, and control structures that are radically different from any existing life form (Blackiston et al., 2021; Ebrahimkhani and Levin, 2021; Kriegman et al., 2020). Because of the tremendous interoperability and plasticity of life, every combination of evolved material (genes, cells, tissues), designed materials, and software is a viable agent and a possible embodiment of sentience (Clawson and Levin, 2022).

Moreover, the field of AI (whether purely software or embodied) is already producing agents with extremely sophisticated performance in linguistic space, while the field of minimal cognition is revealing competences and intelligence in other problem spaces (such as physiological, metabolic, transcriptional, and anatomical). While some of the AIs are able to advocate for themselves verbally (leaving aside the question of whether this behaviour is appropriately grounded or “just faking it”), the bioengineered forms behave in problem spaces we do not normally recognize as intelligent or sentient and do not give verbal reports (this includes cells, organs, and one of your brain hemispheres). It is clear that the human capacity to recognize and evaluate agency is well-tuned for medium sized objects doing interesting things at medium speeds in 3D space, but it is not well adapted to recognizing intelligence and competences in unfamiliar guises and problem spaces (Fields and Levin, 2022).

The coming decades will introduce into society, into our homes, and into our bodies a plethora of novel agents which offer none of the familiar touchstones we have used in the past for gauging moral responsibility toward a given agent: where it came from (factory or evolution) and what it looks like (anatomical structure and homology to humans). Those

simplistic criteria were never appropriate. They were heuristics suitable only for past limitations of imagination and technical capability. It is essential now to develop frameworks that pick out what is *deep and fundamental about sentient beings* – not frozen accidents of evolution on the N=1 example of the phylogenetic path of life on Earth.

3. The future: frameworks for truly diverse intelligences. The Smith & Boyd (1991) criteria are irrelevant across the vast majority of the space of possible agents. Crump et al. use the example of an invertebrate brain (which “differs radically” from our own) to start to stretch these considerations in important ways. Their criterion #7 (associative learning) can occur even in gene regulatory networks (Biswas et al., 2021; Fernando et al., 2009; McGregor et al., 2012; Watson et al., 2010), and most of their other criteria are met by non-neural morphogenetic agents (Friston et al., 2015; Pezzulo and Levin, 2015) via a simple pivot of some of the terms away from “neurons” toward the more general “electrically active cell”. Some of the others (e.g, #8, analgesia preference) form excellent suggestions for future work in non-neural systems and are currently being studied in morphogenetic agents.

1. Nociception
 2. Sensory integration
 3. Integrated nociception
 4. Analgesia: (a) endogenous (b) exogenous
 5. Motivational trade-offs
 6. Flexible self-protection
 7. Associative Learning
 8. Analgesia preference: (a) self-administer (b) location (c) prioritised
- Crump et al.’s (2022) eight criteria (section 2.2)**

Although exo-biology has not yet challenged us with true alien beings, science fiction has been ahead of the game for many decades, warning us about the lack of appropriate frameworks for gauging sentience. The current debate over the sentience of LaMDA (Thopilian et al., 2022) and similar large language models in AI is a case in point. Many have offered very strong opinions about whether LaMDA is or is not sentient, but no one has a good set of criteria that can be used to make such distinctions. If Turing-Test-like evidence is insufficient, so is brain homology. We can’t possibly expect that the only sentient life in the universe has mammalian-style brains; if we can’t agree on intelligence in plants (Calvo et al., 2020; Calvo et al., 2017), and are still wrangling over when and how a “sentience-free” chemistry of the oocyte becomes a human mind, we can have no strong confidence about unconventional agents.

Verbal reports (i.e., language), homology of structure or materials, phylogenetic provenance, etc. are all insufficient to make dependable conclusions across the space of possible agents, some of which will be truly alien in the most crucial ways. And yet, we have an ethical imperative to develop frameworks that recognize sentience beyond our limitations and familiarities. Quantitative criteria like those of Crump et al. are a good example of the kind of framework we need: explicitly laying out conditions in a way that reveals their value and limitations. We will not have certainty, but if we dissolve arbitrary criteria and search for deep invariants across all possible minds and bodies, we will have a morally defensible

position. Maturing to the point of a principled stance on sentience and agency is probably an existential requirement for humankind – in its current embodiment and in the inevitable future ones.

References

- Bakkum, D.J., Chao, Z.C., Gamblen, P., Ben-Ary, G., Shkolnik, A.G., DeMarse, T.B., Potter, S.M., 2007. [Embodying cultured networks with a robotic drawing arm](#). Conference proceedings *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE Engineering in Medicine and Biology Society. Conference 2007, 2996-2999
- Biswas, S., Manicka, S., Hoel, E., Levin, M., 2021. [Gene Regulatory Networks Exhibit Several Kinds of Memory: Quantification of Memory in Biological and Random Transcriptional Networks](#). *iScience* 24, 102131.
- Blackiston, D., Lederer, E.K., Kriegman, S., Garnier, S., Bongard, J., Levin, M., 2021. [A cellular platform for the development of synthetic living machines](#). *Sci Robot* 6, eabf1571.
- Bongard, J., Levin, M., 2021. Living Things Are Not (20th Century) [Machines: Updating Mechanism Metaphors in Light of the Modern Science of Machine Behavior](#). *Frontiers in Ecology and Evolution* 9.
- Calvo, P., Gagliano, M., Souza, G.M., Trewavas, A., 2020. [Plants are intelligent, here's how](#). *Ann Bot* 125, 11-28.
- Calvo, P., Sahi, V.P., Trewavas, A., 2017. [Are plants sentient?](#) *Plant Cell Environ* 40, 2858-2869.
- Chalmers, D., 1996. *The conscious mind*. Oxford University Press, New York.
- Clark, A., Chalmers, D., 1998. [The Extended Mind](#). *Analysis* 58, 7-19.
- Clawson, W.P., Levin, M., 2022. Endless Forms Most Beautiful: teleonomy and the bioengineering of chimeric and synthetic organisms. *Biological Journal of the Linnean Society*, 2022, **XX**, in press.
- Crump, Andrew; Browning, Heather; Schnell, Alex; Burn, Charlotte; and Birch, Jonathan, 2022. [Sentience in decapod crustaceans: A general framework and review of the evidence](#). *Animal Sentience* 32: 1
- Darwin, C., 1859. *On the origin of species by means of natural selection, or, The preservation of favoured races in the struggle for life*. J. Murray, London.
- DeMarse, T.B., Dockendorf, K.P., 2005. [Adaptive flight control with living neuronal networks on microelectrode arrays](#), *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005*. (Vol. 3, pp. 1548-1551).
- Ebrahimkhani, M.R., Levin, M., 2021. [Synthetic living machines: A new window on life](#). *iScience* 24, 102505.
- Fernando, C.T., Liekens, A.M.L., Bingle, L.E.H., Beck, C., Lenser, T., Stekel, D.J., Rowe, J.E., 2009. [Molecular circuits for associative learning in single-celled organisms](#). *Journal of the Royal Society Interface* 6, 463-469.
- Fields, C., Levin, M., 2022. [Competency in Navigating Arbitrary Spaces as an Invariant for Analyzing Cognition in Diverse Embodiments](#). *Entropy* (Basel) 24.

- Friston, K., Levin, M., Sengupta, B., Pezzulo, G., 2015. [Knowing one's place: a free-energy approach to pattern regulation](#). *J R Soc Interface* 12.
- Kagan, B.J., Kitchen, A.C., Tran, N.T., Parker, B.J., Bhat, A., Rollo, B., Razi, A., Friston, K.J., 2021. [In vitro neurons learn and exhibit sentience when embodied in a simulated game-world](#). bioRxiv, 2021.2012.2002.471005.
- Kriegman, S., Blackiston, D., Levin, M., Bongard, J., 2020. [A scalable pipeline for designing reconfigurable organisms](#). *Proc Natl Acad Sci U S A* 117, 1853-1859.
- Levin, M., 2019. [The Computational Boundary of a "Self": Developmental Bioelectricity Drives Multicellularity and Scale-Free Cognition](#). *Front Psychol* 10, 2688.
- Levin, M., 2020. [Life, death, and self: Fundamental questions of primitive cognition viewed through the lens of body plasticity and synthetic organisms](#). *Biochemical and Biophysical Research Communications* 564, 114-133.
- Levin, M., Keijzer, F., Lyon, P., Arendt, D., 2021. [Uncovering cognitive similarities and differences, conservation and innovation](#). *Philos Trans R Soc Lond B Biol Sci* 376, 20200458.
- Lyon, P., 2006. [The biogenic approach to cognition](#). *Cogn Process* 7, 11-29.
- Lyon, P., 2020. [Of what is "minimal cognition" the half-baked version?](#) *Adapt Behav* 28, 407-424.
- Lyon, P., Keijzer, F., Arendt, D., Levin, M., 2021. [Reframing cognition: getting down to biological basics](#). *Philos Trans R Soc Lond B Biol Sci* 376, 20190750.
- McGregor, S., Vasas, V., Husbands, P., Fernando, C., 2012. [Evolution of associative learning in chemical networks](#). *PLoS Computational Biology* 8, e1002739.
- Nagel, T., 1986. *The view from nowhere*. Oxford University Press, New York.
- Orive, G., Taebnia, N., Dolatshahi-Pirouz, A., 2020. [A New Era for Cyborg Science Is Emerging: The Promise of Cyborganic Beings](#). *Adv Healthc Mater* 9, e1901023.
- Pezzulo, G., Levin, M., 2015. [Re-membering the body: applications of computational neuroscience to the top-down control of regeneration of limbs and other complex organs](#). *Integr Biol (Camb)* 7, 1487-1517.
- Pio-Lopez, L., 2021. [The rise of the biocyborg: synthetic biology, artificial chimerism and human enhancement](#). *New Genetics and Society* 40, 599-619.
- Potter, S.M., Wagenaar, D.A., DeMarse, T.B., 2005. [Closing the loop: stimulation feedback systems for embodied MEA cultures](#), in: Taketani, M., Baudry, M. (Eds.), *Advances in Network Electrophysiology Using Multi-Electrode Arrays*. Springer.
- Smith, J.A., Boyd, K.M., Institute of Medical Ethics (Great Britain), 1991. [Lives in the balance : the ethics of using animals in biomedical research : the report of a Working Party of the Institute of Medical Ethics](#). Oxford University Press, Oxford ; New York.
- Thopilan, R., De Freitas, D., Hall, J., et al. (2022) [Lamda: Language models for dialog applications](#). *arXiv preprint arXiv:2201.08239*, 2022.
- Watson, R.A., Buckley, C.L., Mills, R., Davies, A., 2010. [Associative memory in gene regulation networks](#), *Artificial Life Conference XII*, Odense, Denmark, pp. 194-201.