Title: To test the boundaries of consciousness, study animals

Authors:
Simon A. B. Brown (London School of Economics and Political Science)
Elizabeth S. Paul (University of Bristol)
Jonathan Birch (London School of Economics and Political Science)

Correspondence: j.birch2@lse.ac.uk (Jonathan Birch)

Keywords: Distribution of consciousness; sentience; altered states; disorders of consciousness; AI consciousness; Animal Consciousness

Reply to: Tim Bayne, Anil K. Seth, Marcello Massimini, *et al.* (2024). Tests for consciousness in humans and beyond. *Trends in Cognitive Sciences*, https://doi.org/10.1016/j.tics.2024.01.010

Word Count: 774

How can we validate tests for consciousness beyond healthy adult humans? Bayne and colleagues propose the "Iterative Natural Kind Strategy" (INKS) [1]. INKS begins by taking seriously all tests with at least some face validity rather than restricting our attention to a small subset of tests that antecedently pass stringent requirements of validation. We then look for clustering among our tests (and mechanisms explaining such patterns), using these findings to refine our battery of tests. Subsequently, we apply the same process to the new battery.

The strategy raises the question: where to start? Bayne *et al*. make a controversial proposal. They suggest cases can be sorted into "levels" according to "distance" from healthy adult humans, and they propose starting with the "closest" cases first. They note this raises the issue "how populations should be assigned to levels" (p. 9): what is the right concept of "distance"? Are coma patients "nearer" or "further" from healthy human consciousness than non-human animals? Does language render large language models "closer" to us than octopuses—or is language trumped by having a biological brain?

We propose that, given INKS, there are strong reasons for the first phase to include a wide range of non-human animals. Diverse tests have at least some face validity when applied to animals. Furthermore, they can be applied to many different species at different life stages (e.g. juvenile, adult). These features greatly facilitate attempts to study clustering.

Figure 2 of Bayne et al. (2024) gives the opposite impression: that non-human animals (especially invertebrates) are amongst the most “distant” populations, with most tests deemed easier to apply and interpret in humans undergoing disordered states of consciousness, or even in AI. This implies that we should start by investigating AI and coma patients *before* eventually turning our attention to other active, mobile animals with brains. That would be a massive misstep.

The rationales for the claims in Figure 2 are not discussed, and some of these claims seem dubious. For example, why is the perturbation complexity index (PCI) deemed potentially applicable to xenobots and AI, but not to birds? Why are the prospects of the covert command-following test stronger for AI than for monkeys? An excessively skeptical attitude towards consciousness in non-human animals is implicit here.

Strikingly, the paper omits virtually all tests currently used in the animal consciousness literature. Moreover, the one test the authors do include—Unlimited Associative Learning (UAL)—is not a single test. UAL hypothesizes, in short, that a package of advanced learning abilities are enabled by a shared underlying mechanism that suffices for at least a minimal form of consciousness. It is a theory positing a natural kind, not one test.

In reality, the animal consciousness literature is a rich source of tests with at least some face validity ([2], [3]). Current tests can be grouped for convenience as follows (although, importantly, each group contains numerous distinct tests which may ultimately cluster differently [4]): (i) putatively consciousness-linked forms of learning, such as trace conditioning and goal-directed learning; (ii) integration of affective and other information in the service of decision-making; (iii) signs of mental simulation (e.g. vicarious trial and error, model-based reasoning, and tests related to episodic-like memory and planning); (iv) metacognitive and report-like behavior; and (v) indicators of REM-like sleep and dream-like states, including measures related to eye movements and (in octopuses) skin colouration ([5], [6]). Many of these tests can be combined with paradigms that induce dissociations suggestive of distinct conscious and unconscious pathways, with significant differences in performance on tasks due to procedures such as masking and blindsight-related manipulations [7]-[10].

A version of Bayne and colleagues’ table giving all these animal tests their due would look very different (Figure 1).

[Fig. 1 goes here]

A critic might reply: rather than for consciousness *itself*, these may be testing for *other* abilities like integration, learning, or metacognition. However, a parallel concern applies to tests designed for use in clinical settings, such as the PCI: are they really tracking conscious experience *per se* or functional states like coma, sleep, and wakefulness that could occur with or without conscious experience?

INKS, to its credit, urges a pragmatic approach to such questions: do not abandon a test because we cannot be sure from the outset that it tracks consciousness. Instead, iteratively improve our understanding of which tests are tracking consciousness by actively investigating how and why results of our tests cluster. Crucially, this strategy applies equally to clinical tests in humans *and* to the profusion of tests in other animals. Rather than hastily assuming markers are *not* relevant, let us study them, and use the clustering patterns we find to guide iterative theory development.

References

[1] Bayne, T., Seth, A. K., Massimini, M., Shepherd, J., Cleeremans, A., Fleming, S. M., Malach, R., Mattingley, J. B., Menon, D. K., Owen, A. M., Peters, M. A. K., Razi, A., Mudrik, L. (2024). Tests for consciousness in humans and beyond. *Trends in Cognitive Sciences*, https://doi.org/10.1016/j.tics.2024.01.010

[2] Crump, A., Browning, H., Schnell, A., Burn, C., and Birch, J. (2022) Sentience in decapod crustaceans: A general framework and review of the evidence. *Animal Sentience, 32*(1). https://doi.org/10.51291/2377-7478.1691

[3] Paul, E. S., Sher, S., Tamietto, M., Winkielman, P., & Mendl, M. T. (2020). Towards a comparative science of emotion: Affect and consciousness in humans and animals. Neuroscience and biobehavioral reviews, 108, 749–770. https://doi.org/10.1016/j.neubiorev.2019.11.014

[4] Brown, S. A. B., Paul, E. S., & Birch, J. (2024). Potential markers of consciousness in other animals: key references. *Zenodo*. https://doi.org//10.5281/zenodo.11372102

[5] Pophale, A., Shimizu, K., Mano, T. et al. Wake-like skin patterning and neural activity during octopus sleep. Nature 619, 129–134 (2023). https://doi.org/10.1038/s41586-023-06203-4

[6] Anthoney N, Tainton-Heap L, Luong H, Notaras E, Kewin AB, Zhao Q, Perry T, Batterham P, Shaw PJ, van Swinderen B. Experimentally induced active and quiet sleep engage non-overlapping transcriptional programs in Drosophila. Elife. 2023 Nov 1;12:RP88198. https://doi.org/10.7554/eLife.88198

[7] Cowey, A., & Stoerig, P. (1995). Blindsight in monkeys. *Nature*, *373*(6511), 247–249. https://doi.org/10.1038/373247a0

[8] Isa, T., & Yoshida, M. (2021). Neural mechanism of blindsight in a macaque model. *Neuroscience, 469*, 138-161. https://doi.org/10.1016/j.neuroscience.2021.06.022

[9] Droege, P., Weiss, D. J., Schwob, N., & Braithwaite, V. (2021). Trace conditioning as a test for animal consciousness: a new approach. *Animal cognition*, *24*(6), 1299–1304. https://doi.org/10.1007/s10071-021-01522-3

[10] Ben-Haim, M. S., Dal Monte, O., Fagan, N. A., Dunham, Y., Hassin, R. R., Chang, S. W. C., & Santos, L. R. (2021). Disentangling perceptual awareness from nonconscious processing in rhesus monkeys (Macaca mulatta). *Proceedings of the National Academy of Sciences USA*, *118*(15), e2017543118. https://doi.org/10.1073/pnas.2017543118

Fig. 1 (Note to editors: for higher resolution version, see attached file)

		Learning-related tests	Integrative decision-making	Mental simulation	Metacognitive & report-like behaviour	REM-like sleep & dream-like states
(ADULT) NON-HUMAN ANIMALS	Monkeys	+	+	+	+	+
	Rodents	+	+	+	+	+
	Birds	+	+	+	+	+
	Insects	+	+	?	+?	+
	Teleost fish	+	+	+	+?	+
	Decapod crustaceans	+	+	?	?	+
	Cephalopods	+	+	+?	?	+
ALTERED STATES	Sedation	—	?	—	—	+
	Epileptic seizure	—	?	?	?	—
UNCLEAR CAPACITY FOR CONSCIOUSNESS	Disorders of consciousness	—	?	?	—	+
	Babies	?	—	?	—	+
	Fetuses	—	—	—	—	+
ARTIFICIAL SYSTEMS	Neural organoids	?	?	—	—	—
	Xenobots	?	?	?	—	—
	AI	?	?	?	?	—

CAPTION FOR FIG. 1

FIG. 1 TITLE: Impact of considering a wider range of tests.

Revised version of Bayne et al.'s Figure 2, illustrating the impact of emphasizing different tests, and with animals categorized according to biologically meaningful taxa. '+': test can be administered to population (possibly with some modifications), with enough face validity to be studied within the INKS ('+' does *not* imply a test has already been validated); '−': test is either inapplicable or not meaningfully interpretable for the population; '?': more development is needed; '+?': applicable but with unclear interpretation. Each 'test' column corresponds to multiple tests, grouped for convenience. For details and key references, see [4].

Acknowledgements

S.A.B.B. and J.B received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, Grant Number 851145.

E.S.P. was partially funded by BBSRC Grant 'Individual differences in affective processing and implications for animal welfare: a reaction norm approach', Grant reference BB/X014673/1.

Declaration of interests

The authors declare no competing interests.