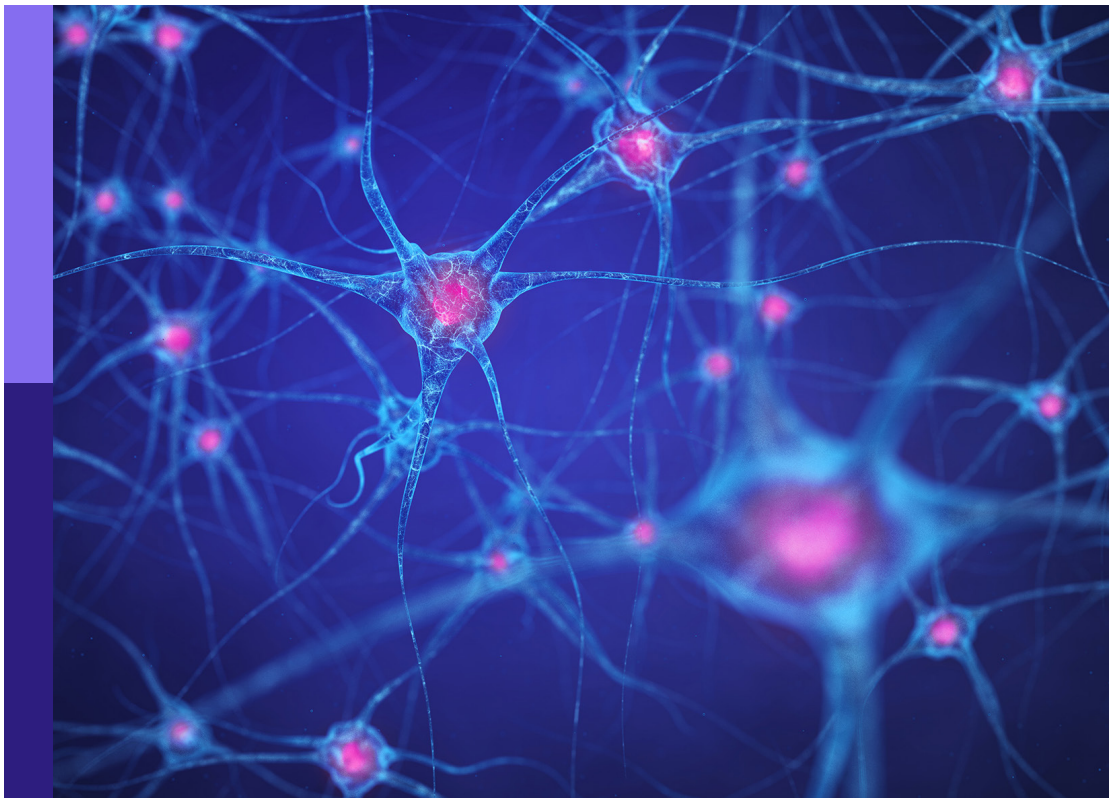
frontiers | Research Topics

Comparative animal consciousness

Edited by
Louis Neal Irwin, Lars Chittka, Nicky S. Clayton,
Eva Jablonka, Jon Mallatt and Todd E. Feinberg

Published in
Frontiers in Systems Neuroscience
Frontiers in Psychology



ISSN 1664-8714
ISBN 978-2-8325-2815-0
DOI 10.3389/978-2-8325-2815-0

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Comparative animal consciousness

Topic editors

Louis Neal Irwin — The University of Texas at El Paso, United States
Lars Chittka — Queen Mary University of London, United Kingdom
Nicky S. Clayton — University of Cambridge, United Kingdom
Eva Jablonka — Tel Aviv University, Israel
Jon Mallatt — Washington State University, United States
Todd E. Feinberg — Icahn School of Medicine at Mount Sinai, United States

Citation

Irwin, L. N., Chittka, L., Clayton, N. S., Jablonka, E., Mallatt, J., Feinberg, T. E., eds. (2023). *Comparative animal consciousness*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-2815-0

Table of contents

frontiers | Frontiers in Systems Neuroscience

TYPE Editorial
PUBLISHED 19 October 2022
DOI 10.3389/fnsys.2022.998421

OPEN ACCESS

EDITED BY
Cyriel Pennartz,
University of Amsterdam, Netherlands

REVIEWED BY
Umberto Olcese,
University of Amsterdam, Netherlands

*CORRESPONDENCE
Louis N. Irwin
lirwin@utep.edu

RECEIVED 20 July 2022
ACCEPTED 10 October 2022
PUBLISHED 19 October 2022

CITATION
Irwin LN, Chittka L, Jablonka E and Mallatt J (2022) Editorial: Comparative animal consciousness.
Front. Syst. Neurosci. 16:998421.
doi: 10.3389/fnsys.2022.998421

Editorial: Comparative animal consciousness

Louis N. Irwin[1]*, Lars Chittka[2], Eva Jablonka[3] and Jon Mallatt[4]

[1]Department of Biological Sciences, The University of Texas at El Paso, El Paso, TX, United States, [2]Research Centre for Psychology, Queen Mary University of London, London, United Kingdom, [3]Cohn Institute for the History of Philosophy of Science and Ideas, Tel Aviv University, Tel Aviv-Yafo, Israel, [4]School of Biological Sciences, Washington State University, Pullman, WA, United States

KEYWORDS

animal cognition, animal awareness, vertebrates, arthropods, cephalopods, definition of consciousness

Editorial on the Research Topic
Comparative animal consciousness

The scientific study of consciousness has seen a resurgence in the 21st century. This collection of reviews, essays, and theories on various aspects of comparative animal consciousness takes a biological and evolutionary approach. As defined here, consciousness refers to the process by which an animal has *perceptual and affective experience or feelings, arising from the material substrate of a nervous system*. It draws upon a long tradition of neuroscientific materialism (Jackson, 1887; Churchland, 1986, 2013; Dennett, 1991; Feinberg, 2012) and a recent emphasis on neurophenomenology (Varela, 1996; Gallagher and Zahavi, 2008; Tononi and Koch, 2015; Irwin and Irwin, 2020; Seth, 2021).

The implications of evolutionary theory for the continuity of life inevitably extended investigations of consciousness to species other than humans. Darwin (1871) believed that consciousness is an evolved capacity, shaped by natural selection and graded in complexity. Arguments for its widespread distribution and ancient origins come from various lines of evidence, including documentation of a variety of different but sufficiently complex, hierarchical neural architectures (Tononi and Edelman, 1998; Dehaene and Naccache, 2001; Dehaene and Changeux, 2011; Barron and Klein, 2016; Feinberg and Mallatt, 2016; Ginsburg and Jablonka, 2019; Carvalho and Damasio, 2021; Chittka, 2022), discovery of sensitivity to stimuli undetectable by humans (Chittka, 2017), behavioral indicators of emotion and self-awareness (Mather, 2008; Baars and Edelman, 2012; Paul et al., 2020; Mallatt et al., 2021; Chittka, 2022), evidence for the adaptive role of associative learning and declarative memory (Bronfman et al., 2016; Ginsburg and Jablonka, 2019), and the cognitive capability for place perception and control of movement (Merker, 2005; Engel, 2010; Chittka and Wilson, 2019; Irwin and Irwin, 2020). Taken together, these studies have led to a growing but not unanimous view that all vertebrates, many arthropods, and cephalopods meet these criteria for sensory and affective consciousness, indicating that consciousness evolved independently in arthropods and vertebrates over half a billion years ago, followed by the cephalopods later in the Paleozoic (Barron and Klein, 2016; Feinberg and Mallatt, 2016; Ginsburg and Jablonka, 2019; Godfrey-Smith, 2020). Alternative theories have been

advanced focusing on mechanisms that likely restrict consciousness to birds, mammals, and some reptiles (Humphrey, 1992; Butler and Cotterill, 2006; Edelman et al., 2011; Pennartz et al., 2019; Nieder et al., 2020) or even to humans alone (Chaisson, 1987; LeDoux, 2019). This collection seeks to shed light on this range of views.

If consciousness arose independently in at least three different clades with very different neural architectures, how many different evolutionary trajectories to consciousness are theoretically possible? The long-standing "multiple realizability thesis" maintains that since so many neural architectures exist across the animal kingdom, the same mental states can arise from an almost unlimited number of different architectures (Putnam, 1967). Mallatt and Feinberg argue that, while different architectures can give rise to different forms of consciousness, the forms that mental states can take are not unconstrained. Since consciousness emerged under the influence of the same vital stimuli (temperature, odors, sounds, electromagnetic waves, etc.) some similarity in neuroanatomy and perceptual content is required in order for different taxa to survive in competition in the same physical world. At the same time, others emphasize that as evolutionary pressures differ profoundly between species, so do their sense organs, perceptual systems, and mental operations (Bräuer et al., 2020; Montemayor, 2021; Chittka, 2022).

Thurston Lacalli reasons that consciousness evolved like all biological attributes, from simple antecedents that were progressively elaborated and refined over an extended period of evolutionary time as stepwise adaptations in different cognitive niches. In his first contribution to this volume, Lacalli focuses on "selector circuits" of neurons that encode irreducible elements of experience (qualia) as subunits of the neural correlates of consciousness that evolve through progressive refinement. In his second contribution, Lacalli envisions how distinctive qualia evolved from more diffuse and less differentiated "original (ur-) qualia."

Two articles on how natural selection can channel consciousness toward greater complexity are included here. Tjøstheim et al. note that navigation, including taking detours, appears to be an essential element of consciousness, because it requires map-like cognitive structures for spatial representation beyond the animal's immediate location. By using simulations in a forced detour paradigm, they show how different strategies can yield behaviors that approximate those of different species. They propose that both neuronal population size and inhibitive efficacy may be important for allowing organisms to negotiate predation risks and natural geometries that obstruct foraging.

In the second example,Van De Pol and van Swinderen build on the paradoxical view of brain function as an ongoing balance between prediction and surprise as a factor in understanding the evolution of consciousness. In particular, this view may provide insight into the function and evolution of active sleep, which is widespread in animals, not just in mammals and birds. They suggest that such sleep evolved as a mechanism for refining and generalizing internal models of the world during sleep, to minimize prediction errors in the waking state.

The earliest vertebrates to evolve were jawless fishes. Suzuki reviews the evidence that the surviving members of that clade — lampreys and hagfish — display the markers of primary, minimal consciousness. He concludes that the adult lamprey appears to meet the neuroanatomical criteria for mediating consciousness. While less is known about hagfish, their sensory behaviors and learning abilities are more amenable to lab testing, and may soon provide the basis for conclusions about their capacity for consciousness as well.

Molluscs, with mostly small brains or merely dispersed ganglia, separated from the lineage to vertebrates over 550 million years ago. But cephalopods soon diverged as a molluscan subgroup, evolving large nervous systems, complex behavior, and significant cognitive abilities (Young, 1964; Grasso and Basil, 2009; Schnell and Clayton, 2021). In a wide-ranging review of historical and current research on the neuroanatomy, behavior, and cognitive abilities of cephalopods, Ponte et al. conclude from five different criteria that these animals have the capacity for at least a basal faculty of consciousness. They further advocate for asking, not "Is this species more conscious than that one?" but rather, "How is the individual experience of this species different from that one?" Kaufmann endorses that formulation, pointing to the growing realization that placing an organism on a single sliding-scale model for consciousness is a methodological mistake. Rather, the behavioral, cognitive and neurological criteria for conscious experience should be sensitive to experience-specific differences conceived within a multidimensional framework that provides a distinct consciousness profile for each species. An example of a non-linear multidimensional model gaining traction is one proposed by Birch et al. (2020).

The paper by Carls-Diamante questions the common notion that consciousness must have a unified structure by noting that a majority of the neurons in an octopus are found, not in its brain, but in its arms. She raises the intriguing idea that each octopus arm may be capable of supporting its own idiosyncratic field of consciousness, limited in content to the sensory and motor processes relevant to that arm. She then points out that if we are to have a more comprehensive understanding of different types of creature consciousness, particularly among invertebrates, we need to go beyond vertebrate-based assumptions about phenomenal experience, such as the notions that there is only one conscious field per organism and that only the CNS can generate conscious fields.

Numerous authors have viewed motility as a primary driver for the evolution of consciousness (Sheets-Johnstone, 1999; Merker, 2005; Engel, 2010; Chittka and Wilson, 2019). Vallortigara likewise makes the core assumption that animals have evolved phenomenal experience in strict association with active movement. Here and in previous writing (Vallortigara,

2020), he invokes the concept of the internally-generated efference copy to distinguish between sensations (what is happening to me, internally) from perceptions (what is happening out there, externally), as originally proposed by Humphrey (1992). Vallortigara argues that consciousness arises from the interplay of this internally-generated efference copy and sensory input from the outside world.

Problem solving through insight may be another window into animal consciousness. Though difficult to investigate in non-verbal animals, Osuna-Mascaro and Auersperg suggest that it may be widespread and amenable to study through proxy indicators, such as eye-tracking, pupil dilation, and emotionality.

Michael Levin provides an overarching perspective that places animal consciousness as a process within a broader population of "cognitive systems," and invites a reconsideration of the traditional limited conceptions of cognition, the self, memory, regeneration, developmental programs, and evolution. His article provides many novel insights, including questions of agency, the nature of the Self, an expansive view of intelligence, the operation and architecture of distributed memory, and various aspects of consciousness.

The net effect of the contributions to this volume is to support the growing acceptance of the idea that consciousness is ancient in origin and widespread across the phylogenetic spectrum, arising in a diversity of nervous systems, and manifested in a variety of ways (Darwin, 1871; Koch, 2012; Feinberg and Mallatt, 2016, 2018; Chittka and Wilson, 2019; Ginsburg and Jablonka, 2019; Irwin, 2020). They also point to the need for a definition of consciousness, like the one proposed in the first paragraph of this editorial, that is generic enough to encompass a broad range of animal phenomenologies.

Author contributions

LI wrote the draft version of the paper and oversaw revisions. LC, EJ, and JM made extensive suggestions of both an editorial and substantive nature. All authors have read and approved the final version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Baars, B. J., and Edelman, D. B. (2012). Consciousness, biology and quantum hypotheses. *Phys. Life Rev.* 9, 285-294. doi: 10.1016/j.plrev.2012.07.001

Barron, A. B., and Klein, C. (2016). What insects can tell us about the origins of consciousness. *Proc. Natl. Acad. Sci. U S A.* 113, 4900–4908. doi: 10.1073/pnas.1520084113

Birch, J., Schnell, A. K., and Clayton, N. S. (2020). Dimensions of animal consciousness. *Trends Cogn Sci.* 24, 789–801. doi: 10.1016/j.tics.2020.07.007

Bräuer, J., Hanus, D., Pika, S., Gray, R., and Uomini, N. (2020). Old and new approaches to animal cognition: There is not "one cognition". *J. Intell.* 8, 28. doi: 10.3390/jintelligence8030028

Bronfman, Z. Z., Ginsburg, S., and Jablonka, E. (2016). The transition to minimal consciousness through the evolution of associative learning. *Front. Psychol.* 7, 1954. doi: 10.3389/fpsyg.2016.01954

Butler, A. B., and Cotterill, R. M. (2006). Mammalian and avian neuroanatomy and the question of consciousness in birds. *Biol. Bull.* 211, 106–127. doi: 10.2307/4134586

Carvalho, G. B., and Damasio, A. (2021). Interoception and the origin of feelings: A new synthesis. *Bioessays.* 43, e2000261. doi: 10.1002/bies.202000261

Chaisson, E. (1987). *The Life Era: Cosmic Selection and Conscious Evolution*. Boston, MA: Atlantic Monthly Press.

Chittka, L. (2017). Bee cognition. *Curr. Biol.* 27, R1049–R1053. doi: 10.1016/j.cub.2017.08.008

Chittka, L. (2022). *The Mind of a Bee*. Princeton, NJ: Princeton University Press.

Chittka, L., and Wilson, C. (2019). Expanding consciousness. *Amer. Scientist* 107, 364–369. doi: 10.1511/2019.107.6.364

Churchland, P. M. (2013). *Matter and Consciousness (3rd ed.)*. Cambridge, MA: MIT Press.

Churchland, P. S. (1986). *Neurophilosophy*. Cambridge: MIT Press.

Darwin, C. (1871). *The Descent of Man (2nd ed.)*. New York: A. L. Burt.

Dehaene, S., and Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron* 70, 200–227. doi: 10.1016/j.neuron.2011.03.018

Dehaene, S., and Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition* 79, 1–37. doi: 10.1016/S0010-0277(00)00123-2

Dennett, D. C. (1991). *Consciousness Explained*. Boston: Little, Brown and Co.

Edelman, G. M., Gally, J. A., and Baars, B. J. (2011). Biology of consciousness. *Front. Psychol.* 2, 4. doi: 10.3389/fpsyg.2011.00004

Engel, A. K. (2010). "Directive minds: How dynamics shapes cognition," in J. Stewart, O. Gapenne and E. A. Di Paolo (Eds.), *Enaction: Toward a New Paradigm for Cognitive Science* (Cambridge, MA: MIT Press) 219–243. doi: 10.7551/mitpress/9780262014601.003.0009

Feinberg, T. E. (2012). Neuroontology, neurobiological naturalism, and consciousness: a challenge to scientific reduction and a solution. *Phys. Life Rev.* 9, 13–34. doi: 10.1016/j.plrev.2011.10.019

Feinberg, T. E., and Mallatt, J. M. (2016). *The Ancient Origins of Consciousness: How the Brain Created Experience*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/10714.001.0001

Feinberg, T. E., and Mallatt, J. M. (2018). *Consciousness Demystified*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/11793.001.0001

Gallagher, S., and Zahavi, D. (2008). *The Phenomenological Mind: An Introduction to Philosophy of Mind and Cognitive Science (1st ed.)*. New York: Routledge. doi: 10.4324/9780429319792-1

Ginsburg, S., and Jablonka, E. (2019). *The Evolution of the Sensitive Soul: Learning and the Origins of Consciousness* Cambridge, MA: MIT Press. doi: 10.7551/mitpress/11006.001.0001

Godfrey-Smith, P. (2020). *Metazoa: Animal Minds and the Birth of Consciousness*. London: William Collins.

Grasso, F. W., and Basil, J. A. (2009). The evolution of flexible behavioral repertoires in cephalopod molluscs. *Brain Behav. Evol.* 74, 231–245. doi: 10.1159/000258669

Humphrey, N. (1992). *A History of the Mind: Evolution and the Birth of Consciousness*. New York: Copernicus - Springer-Verlag. doi: 10.1007/978-1-4419-8544-6

Irwin, L. N. (2020). Renewed perspectives on the deep roots and broad distribution of animal consciousness. *Front. Syst. Neurosci.* 14, 57. doi: 10.3389/fnsys.2020.00057

Irwin, L. N., and Irwin, B. A. (2020). Place and environment in the ongoing evolution of cognitive neuroscience. *J. Cogn. Neurosci.* 32, 1837–1850. doi: 10.1162/jocn_a_01607

Jackson, J. H. (1887). Remarks on evolution and dissolution of the nervous system. *J. Mental Sci.* 23, 25–48. doi: 10.1192/bjp.33.141.25

Koch, C. (2012). *Consciousness: Confessions of a Romantic Reductionist*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/9367.001.0001

LeDoux, J. (2019). *The Deep History of Ourselves: The Four-Billion-Year Story of How We Got Conscious Brains*. Viking, New York.

Mallatt, J., Blatt, M. R., Draguhn, A., Robinson, D. G., and Taiz, L. (2021). Debunking a myth: plant consciousness. *Protoplasma.* 258, 459–76. doi: 10.1007/s00709-020-01579-w

Mather, J. A. (2008). Cephalopod consciousness: behavioural evidence. *Conscious Cogn.* 17, 37–48. doi: 10.1016/j.concog.2006.11.006

Merker, B. (2005). The liabilities of mobility: a selection pressure for the transition to consciousness in animal evolution. *Conscious Cogn.* 14, 89–114. doi: 10.1016/S1053-8100(03)00002-3

Montemayor, C. (2021). Types of consciousness: The diversity problem. *Front. Syst. Neurosci.* 15, 747797. doi: 10.3389/fnsys.2021.747797

Nieder, A., Wagener, L., and Rinnert, P. (2020). A neural correlate of sensory consciousness in a corvid bird. *Science* 369, 1626–1629. doi: 10.1126/science.abb1447

Paul, E. S., Sher, S., Tamietto, M., Winkielman, P., and Mendl, M. T. (2020). Towards a comparative science of emotion: Affect and consciousness in humans and animals. *Neurosci. Biobehav. Rev.* 108, 749–770. doi: 10.1016/j.neubiorev.2019.11.014

Pennartz, C. M. A., Farisco, M., and Evers, K. (2019). Indicators and criteria of consciousness in animals and intelligent machines: An inside-out approach. *Front. Syst. Neurosci.* 13, 25. doi: 10.3389/fnsys.2019.00025

Putnam, H. (1967). Psychological predicates. *Art Mind Religion* 1, 37–48.

Schnell, A. K., and Clayton, N. S. (2021). Cephalopods: Ambassadors for rethinking cognition. *Biochem. Biophys. Res. Commun.* 564, 27–36. doi: 10.1016/j.bbrc.2020.12.062

Seth, A. K. (2021). *Being You: A New Science of Consciousness*. New York: Penguin Random House.

Sheets-Johnstone, M. (1999). *The Primacy of Movement*. Amsterdam: John Benjamins Publishing. Vol. 14. doi: 10.1075/aicr.14

Tononi, G., and Edelman, G. M. (1998). Consciousness and complexity. *Science* 282, 1846–1851. doi: 10.1126/science.282.5395.1846

Tononi, G., and Koch, C. (2015). Consciousness: here, there and everywhere? *Philos. Trans. R Soc. Lond B Biol. Sci.* 370, 20140167. doi: 10.1098/rstb.2014.0167

Vallortigara, G. (2020). The rose and the fly. A conjecture on the origin of consciousness. *Biochem. Biophys. Res. Commun.* 564, 170–174. doi: 10.1016/j.bbrc.2020.11.005

Varela, F. (1996). Neurophenomenology: A methodological remedy for the hard problem. *J. Consciousness Stud.* 3, 330–349.

Young, J. (1964). *A Model of the Brain*. London: Oxford Univ Press.

frontiers in Systems Neuroscience

PERSPECTIVE
published: 10 September 2021
doi: 10.3389/fnsys.2021.741579

Experience-Specific Dimensions of Consciousness (Observable in Flexible and Spontaneous Action Planning Among Animals)

*Angelica Kaufmann**

Cognition in Action Unit, PhiLab, University of Milan, Milan, Italy

The multidimensional framework to the study of consciousness, which comes as an alternative to a single sliding scale model, offers a set of experimental paradigms for investigating dimensions of animal consciousness, acknowledging the compelling urge for a novel approach. One of these dimensions investigates whether non-human animals can flexibly and spontaneously plan for a future event, and for future desires, without relying on reinforcement learning. This is a critical question since different intentional structures for action in non-human animals are described as served by different neural mechanisms underpinning the capacity to represent temporal properties. And a lack of appreciation of this variety of intentional structures and neural correlates has led many experts to doubt that animals have access to temporal reasoning and to not recognize temporality as a mark of consciousness, and as a psychological resource for their life. With respect to this, there is a significant body of ethological evidence for planning abilities in non-human animals, too often overlooked, and that instead should be taken into serious account. This could contribute to assigning consciousness profiles, across and within species, that should be tailored according to an implemented and expansive use of the multidimensional framework. This cannot be fully operational in the absence of an additional tag to its dimensions of variations: the *experience-specificity* of consciousness.

Keywords: animal consciousness, action plan, temporal cognition, ethology, comparative psychology

OPEN ACCESS

Edited by:
Louis Neal Irwin,
The University of Texas at El Paso,
United States

Reviewed by:
Akane Nagano,
Kyoto University, Japan
Rocco J. Gennaro,
University of Southern Indiana,
United States

***Correspondence:**
Angelica Kaufmann
angelica.kaufmann@gmail.com

Received: *14 July 2021*
Accepted: *25 August 2021*
Published: *10 September 2021*

Citation:
Kaufmann A (2021) Experience-Specific Dimensions of Consciousness (Observable in Flexible and Spontaneous Action Planning Among Animals). Front. Syst. Neurosci. 15:741579. doi: 10.3389/fnsys.2021.741579

INTRODUCTION

Cognition varies extensively in nature as individuals adapt to the specific challenges they experience in life (Irwin, 2020). Sumatran and Bornean orangutans, for example, have developed impressive vocal communicative skills because they live in isolation in a very dense arboreal environment in which individuals of a population cannot rely on a visually transmissible communicative repertoire, like gestures. On the contrary, chimpanzees in Uganda and bonobos in DR Congo, do not live in isolation and have developed sophisticated gestural repertoires that they use to

communicate. Boesch's (2021) calls this *experience-specific cognition*. Such considerations over cognition ought to be extended to the study of comparative animal consciousness, a field of research that could be accordingly rebranded as *experience-specific consciousness*.

The recent multidimensional framework to the study of consciousness (Birch et al., 2020), which comes as an alternative to a single sliding scale model, offers a set of experimental paradigms for investigating dimensions of animal consciousness, acknowledging the compelling urge for a novel approach. One of these dimensions investigates whether non-human animals can flexibly and spontaneously plan for a future event, and for future desires, without relying on reinforcement learning. This is a critical question since different intentional structures for action in non-human animals are described as served by different neural mechanisms underpinning the capacity to represent temporal properties (Cai et al., 2012; Mayo and Sommer, 2013; Schormans et al., 2017; Feenders and Klump, 2018; Perry and Chittka, 2019; Viera and Margolis, 2019). And a lack of appreciation of this variety of intentional structures and neural correlates has led experts (Hoerl and McCormack, 2019; Redshaw and Suddendorf, 2020) to doubt that animals can have access to temporal reasoning and to not recognize temporality as a mark of consciousness, and as a psychological resource for their life. With respect to this, there is a significant body of ethological evidence for planning abilities in non-human animals, too often overlooked, and that instead should be taken into serious account. This could contribute to assigning consciousness profiles, across and within species, that should be tailored according to an implemented and expansive use of the multidimensional framework. This cannot be fully operational in the absence of an additional tag to its dimensions of variations: the *experience-specificity* of consciousness.

To appreciate the significant change of perspective that is now encouraging researchers to treat the subject of consciousness in novel terms, it shall be noticed that since not very long ago, consciousness was one of those subjects that researchers were advised not to write about up until tenure. Even now, if one writes about consciousness in non-human animals that person should be ready to face the dubious looks from a lot of skeptics (see Andrews, 2016; Allen and Trestman, 2017, 2020, for a review of arguments for and against animal consciousness from a philosophical and empirical perspective). But the wall of skepticisms toward the legitimacy to write about consciousness, and especially about consciousness in non-human animals began to fall with, courtesy of the Cambridge Declaration on Consciousness (Low et al., 2012). This document assesses that the neurological substrates of all mammals, birds, and many other creatures, including octopuses, are complex enough to support conscious experience. As a result, the first achievement of this change of perspective was the fact that the question was no longer as to whether animals other than humans were conscious, but what their consciousness would look like. The second significant and unprecedented achievement since 2012, was that of seeing researchers acknowledging that to place an organism on a single sliding scale model for consciousness at the top of which—that goes without saying—we would find humans, is a methodological mistake, symptomatic of a widespread tendency resulting from a failure to meet the two following explanatory targets: no-underestimation principle and no-overascription principle. The first one is the principle according to which we should not underestimate the richness of all animal experiences since the neurological substrates for conscious experience are present in a variety of forms among non-human animals. And the second one is the principle according to which we should not overascribe supposedly desirable similarities between non-human animal experience and human animal experience since the neurological substrates of human conscious experience are one among various different neurological structures allowing for conscious experience.

A recent proposal presents itself as an excellent candidate to meet both principles. This is the multidimensional framework to the study of consciousness presented by Birch et al.'s (2020) work, which outlines a set of experimental paradigms for investigating dimensions of animal consciousness, as an alternative to a single sliding scale model. They highlight five significant dimensions of variation within and across animal species: perceptual richness (p-richness), evaluative richness (e-richness), integration at a time (unity), integration across time (temporality), and self-consciousness (selfhood). Taking the case of integration across time will allow for the introduction of an additional tag to these dimensions of variations: the *experience-specificity* of consciousness.

CONSCIOUSNESS IS INTEGRATION ACROSS TIME

Various researchers (Osvath and Martin-Ordas, 2014; Müller et al., 2017; Martin-Ordas, 2020; Martin-Ordas et al., 2020; van Leeuwen, 2021) have contributed evidence on the relationship between the experience of time and agency in the specific experience of non-human animals that supports the proposal advanced by Birch, Schnell and Clayton that a multidimensional framework is beneficial to the study of consciousness within the same animal species and across different animal species.

To discuss the relationship between temporal experience and agency, the present focus is on integration across time (temporality), and especially on future planning. When we act, we act across time, and human beings along with many other species, are capable of producing and expressing complex intentional structures for action (Dickinson, 2012). Behavioral manifestations of such complex structures suggest that various creatures possess temporal understanding (Hoerl and McCormack, 2001), but that they cannot reason about time (Hoerl and McCormack, 2019). That is, non-human animals seem able to represent temporal properties such as duration, order of events, causal links between events, and to represent time as passing, but they lack the capacity to understand time as a measure of change (see for example, Blaisdell et al., 2006). However, representing time as a measure of change is an essential aspect of action planning, and thereby providing an account of how different animal species represent time according to their

specific experiences is a crucial component in any investigation of their capacity for action planning (Kaufmann, 2015, 2016; Safina, 2016; Kaufmann and Cahen, 2019).

van Schaik et al. (2013) argue that the capacity for action planning relies on two cognitive abilities: self-control and mental time travel. Self-control is understood as the capacity to repress one's own immediate need and postpone a reward (Osvath and Osvath, 2008; MacLean et al., 2014). Mental time travel is defined as the capacity to mentally represent potential future events (Clayton et al., 2003; Tulving, 2005; Rosati et al., 2007; Roberts and Feeney, 2009; Corballis, 2019). These two core skills that a cognitive system needs to plan future actions are, arguably, complementary. Evidence shows that the capacity that many non-human animals have for mental time travel is at play in a variety of planned actions, such as tool-using practices and anticipatory vocalizations, among other cases (Osvath et al., 2012).

We will look at tool-using first. Chimpanzees can appreciate the difference between present and future uses of the same tool, and they can articulate a coherent sequence of time-displaced intentional actions that involve that object. Since the vast majority of empirical evidence for tools manipulation over time concerns stones, these activities are grouped under the label of "stone handling" behaviors (Cenni et al., 2020). The empirical literature on the matter is flourishing (Bobrowicz et al., 2020). We benefit from various reports on goal-oriented anticipatory behavior like termite fishing and nut-cracking (Boesch and Boesch, 1990; Voelter and Call, 2014), moss-sponging and leaf-sponge re-use (Hobaiter et al., 2014). It is still a matter of controversy whether we can infer instances of action planning from these studies. One reason is that these studies were not meant to investigate planning capacities directly.

The first study that directly addressed a question on action planning capacities, and that provided positive results, shows that orangutans and bonobos can save tools for future use (Mulcahy and Call, 2006); a second study discovered the same abilities in orangutans and chimpanzees as well (Osvath and Osvath, 2008); a third study reinforced these findings with new evidence on chimpanzees (Dufour and Sterck, 2008); and a fourth, but indeed the first agreed upon piece of unambiguous evidence of planning capacities in non-human primates is that recorded by Osvath (2009). This study focused on a captive male chimpanzee (*Pan troglodytes*), named Santino, who was observed (for over 10 years) to have very articulated dominance displays: hurling stones at zoo visitors. The animal would intentionally select, store, conceal, and eventually throw stones at others with the intent of showing dominance. His behavior did not go unnoticed, because even after the zookeepers had cleaned up the compound from every stone, Santino would manage to continue hurling other stones. He started to collect stones from the water moat that surrounds the outside compound. Santino stored them for a later purpose. The chimpanzee behavior has been thus analyzed: the first phase includes the selection, collection, and concealment of the stones. The second phase consists in the manufacturing of discs from concrete, when ready at hand stones were not available. The third phase is the use of these objects as weapons to hurl at zoo visitors. In Osvath and Karvonen (2012), they improved the experimental procedure of their observational studies and reported what follows: the manufactures from concealment become the preferred weapon. The chimpanzee positioned these concealments very close to the visitors' observation area. He started to deploy a two-step deceptive strategy: firstly, the chimpanzee kept his "weapons" occluded from the visitors' visual space (see Hare et al., 2001, for evidence that chimpanzees appreciate when something in their visual field is unavailable to someone else's sight), and secondly, he inhibited his dominance display behavior in order not to scare the visitors and keep them close enough to the observation area. Notably, the chimpanzee had a calm attitude during the first and the second phase, while he got very agitated during the third one—as if he could appreciate the fact that showing arousal from the beginning was going to scare the onlookers ahead of time and compromise the plan.

Osvath classifies this behavior as a planned activity because it is a time-structured intentional action that can be further divided into sub-phases or sub-plans. Santino intends to display dominance, and his plan is a threefold activity extended to the future. Osvath maintains that: "In order for a behavior to signal planning for a future state the predominant mental state during the planning must deviate from the one experienced in the situation that is planned for. The above behavior is clearly identifiable as planning for a future state" (Osvath, 2009, p. 191). The *predominant mental state* is the intentional structure that triggers and subsequently guides the plan throughout its phases. As such, intentions deviate from the mental states that guide the ongoing planned activity at the time it is being experienced. In addition to the threefold structure of the stone hurling planned activity, there are two distinct behaviors to be highlighted: firstly, the chimpanzee's ability to appreciate whether a given object falls within or outside of the visual field and space of action of a potentially competing third part, and how this affects the structure of the plan of action; secondly, the chimpanzee's awareness of the fact that repressing its own dominant attitude could bring an advantage toward the achievement of the intended outcome. These two behaviors exemplify the capacity for cross-temporally referential connectivity, individuated by Bratman (1987, 2014), that is the feature of intentions that characterizes these mental states as both backward and forward-looking.

A different observational study by van Schaik et al. (2013) examined the extent to which the direction of long calls emitted by male Sumatran orangutans (*Pongo abelii*) and Bornean orangutans (*Pongo pygmaeus wurmbii*) indicated the direction of their future travel. These animals live in a very dense tropical forest and are semi-solitary, thus often out of sight from other members of their population. The goal of male orangutan's long calls is that of indicating to female members the future travel direction of the male. Vocalizations are performed by individuals when stationary and can anticipate the direction of their travel 1 day ahead. The study of van Schaik et al. (2013) focused on three issues: first, they tested whether the direction in which flanged male Sumatran orangutans give spontaneous long calls generally predicts the subsequent travel direction. Second, they investigated whether a new spontaneous long call indicates the subsequent travel direction better than the old

one would have if no new call had been given. Third, they tested the extent to which long calls given in the evening at or near the night nest still indicate travel direction during the next day, thus indicating future planning independent of the current motivational state. The temporal dimension of consciousness is particularly interesting with respect to the evidence at hand about the capacity displayed by male Sumatran orangutans and Bornean orangutans to communicate their future travel directions and the corresponding ability displayed by female orangutans to be receptive to such communicative intentions (van Schaik et al., 2013; Spillman et al., 2015; Askew and Morrogh-Bernard, 2016; Lameira and Call, 2018). As described, together tool-use and travel calls provide fertile ground for discussing integration across time as a marking dimension for a consciousness profile. Yet, this type of evidence is not properly acknowledged within the multidimensional framework.

INTEGRATION ACROSS TIME IS BEST OBSERVED IN FLEXIBLE AND SPONTANEOUS BEHAVIOR

The multidimensional framework and its current experimental paradigms can be informed by implementing the empirical literature, currently deployed, with more evidence from ethology, in addition to evidence from comparative experimental psychology. In particular, as said, this analysis focuses on evidence that emphasizes the presence in non-human primates of the capacity for integration across time and temporal reasoning. To explain why ethology matters in this context, I shall discuss this dimension of consciousness in terms of the Lean Temporal Integration Approach and Rich Temporal Integration Approach. The multidimensional framework buys elements of both approaches, reasonably so. The first and fundamental difference between the two is given by methodology. The Lean Temporal Integration Approach is built on the research methods of comparative experimental psychology, that is, behavioral experiments run in artificial settings (Tomasello and Call, 1997; Leavens et al., 2010; Webster and Rutz, 2020); the Rich Temporal Integration Approach is the result of the research methods of cognitive ethology, that is, research in the field, mostly done as observation of animal behavior in the wild (Nishida et al., 1983; Healy et al., 2009; Smulders et al., 2010; Janmaat et al., 2014, 2016; Rosati, 2017; Boesch, 2020, 2021; Bräuer et al., 2020). These two approaches lead to very different conclusions about the structure of non-human animal experience: the Lean Integration Approach argues for a lack of motivation in pursuing action planning on the side of the animal, and from this lack of motivation it infers a lack of cognitive faculties that are needed to act spontaneously toward a future goal. Conversely, the Rich Integration Approach distinguishes evidence for lack of motivation to interpretations about lack of cognitive capacities. When the comparative experimental psychologist asks the question of what a certain species is capable of achieving in terms of spontaneous future goals, she is investigating the motivational aspect of instances that can reflect this behavior. When the ethologist asks the question of what is possible to achieve in terms of spontaneous future goals, she is investigating the behavioral criteria that can account for this cluster of flexible action plans. I argue that the two claims of the Lean Temporal Integration Approach can and ought to be kept separate: evidence that non-human animals are mostly pursuing repetitive activities motivated by recurrent goals is not evidence that they are only capable of pursuing recurrent goals. Evidence that non-human animals appreciate the recurrent nature of the goals of others is not evidence that they are capable only of ascribing recurrent goals to others. I concede to the Lean Temporal Integration Approach that the vast majority of non-human animals activities is driven by recurrent goals and by the capacity to ascribe recurrent goals to others; what I disagree with, in the context of the Lean Temporal Integration Approach is the assumption that this capacity to form and ascribe recurrent goals is limited to recurrent goals. Evidence from empirical research in support of the Rich Temporal Integration Approach points to the fact that non-human animals are capable of forming and ascribing spontaneous and flexible goals that extend to articulated actions. The purpose of presenting these two approaches is to highlight the fact that the analysis of the five dimensions of variation should be sensitive as to whether the evidence taken into account at a time is obtained from observational work or from a controlled environment. A consciousness profile of a given animal species drawn from evidence from ethology, would in all likelihood differ from one tailored from evidence from comparative experimental psychology.

I have exemplified this methodological difference between the two approaches by focusing on specific observational studies. Out of the various empirical evidence ascribing consciousness to non-human animals, I turned attention to evidence from ethology, which are revealing of the richness of animal cognition, crucial to consciousness and made manifest by the spontaneity and flexibility of action (Pennartz et al., 2019). I wanted to explain how, through a Rich Temporal Integration Approach, consciousness can be observed in various species and how non-human animals can be assigned a consciousness profile tailored according to the specificity of their experience.

CONCLUSION

The analysis of the dimensions of variation should be sensitive to whether the evidence taken into account at a time is obtained from observational work or from a controlled environment. As explained, a multidimensional framework and its current experimental paradigms can be informed by implementing the empirical literature with more evidence from ethology, in addition to evidence from comparative experimental psychology.

Conscious experience is assessed through a series of behavioral, cognitive and neurological criteria. Firstly, contrary to what most people assumed until a decade ago, the impossibility of collecting verbal reports from animals does not preclude the scientific investigation of animal consciousness. It is not only animals that are incapable of providing verbal reports about their inner life, but also young children and patients in minimally

conscious states. And since most people will not deny conscious experience to children or such patients, so they should not deny conscious experience to other animals. Secondly, to deploy a single sliding scale model for measuring consciousness, would amount to following a fallacious methodology and a hardly scientific one, not least, as just said, because conscious experience cannot and should not be investigated according to rigid criteria such as verbal reports. For these reasons, the behavioral, cognitive and neurological criteria for conscious experience should be sensitive to dimensions of variation that should exist within a multidimensional framework conceived in order to provide a different consciousness profile for each animal species.

In particular, as discussed, consciousness can be observed in the flexible and spontaneous planning behavior of various primates, and these animals can be given a consciousness profile tailored according to an implemented and expansive use of the multidimensional framework which ought to take into account an additional tag to its dimensions of variations: the *experience-specificity* of consciousness.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

ACKNOWLEDGMENTS

I thank the animals mentioned for tolerating human presence over so many years and for showing humans how they solve the many fascinating challenges of their life, making clear to us the importance of population and cultural differences.

REFERENCES

Allen, C., and Trestman, M. (2017). "Animal consciousness," in *The Blackwell Companion to Consciousness*, eds S. Schneider and M. Velmans (Hoboken, NJ: John Wiley & Sons, Ltd), doi: 10.1002/9781119132363.ch5

Allen, C., and Trestman, M. (2020). *Animal Consciousness*. The Stanford Encyclopedia of Philosophy, Available online at: https://plato.stanford.edu/archives/win2020/entries/consciousness-animal/. (Accessed October 24, 2016).

Andrews, K. (2016). *Animal Cognition*. The Stanford Encyclopedia of Philosophy, Availableonline at: http://plato.stanford.edu/archives/sum2016/entries/cognition-animal/. (Accessed May 6, 2016)

Askew, J. A., and Morrogh-Bernard, H. C. (2016). Acoustic characteristics of long calls produced by male orang-Utans (Pongo pygmaeus wurmbii): advertising individual identity, context, and travel direction. *Folia Primatol.* 87, 305–319. doi: 10.1159/000452304

Birch, J., Schnell, A. K., and Clayton, N. S. (2020). Dimensions of animal consciousness. *Trends Cogn. Sci.* 24, 789–801. doi: 10.1016/j.tics.2020.07.007

Blaisdell, A., Sawa, K., Leising, K. J., and Waldmann, M. R. (2006). Causal reasoning in rats. *Science* 311, 1020–1022. doi: 10.1126/science.1121872

Bobrowicz, K., Johansson, M., and Osvath, M. (2020). Great apes selectively retrieve relevant memories to guide action. *Sci. Rep.* 10:12603. doi: 10.1038/s41598-020-69607-6

Boesch, C. (2020). Mothers, environment, and ontogeny affect cognition. *Anim. Behav. Cogn.* 7, 747–489.

Boesch, C. (2021). Identifying animal complex cognition requires natural complexity. *iScience* 24:3. doi: 10.1016/j.isci.2021.102195

Boesch, C., and Boesch, H. (1990). Tool usen and tool making in wild chimpanzees. *Folia Primatol.* 54, 86–99. doi: 10.1159/000156428

Bratman, M. (1987). *Intention, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press.

Bratman, M. (2014). *Shared Agency: a Planning Theory of Acting Together*. Oxford: Oxford University Press.

Bräuer, J., Hanus, D., Pika, S., Gray, R., and Uomini, N. (2020). Old and new approaches to animal cognition: there is not 'One Cognition'. *J. Intell.* 8:28. doi: 10.3390/jintelligence8030028

Cai, M., Stetson, C., and Eagleman, D. M. (2012). A neural model for temporal order judgments and their active recalibration: a common mechanisms for space and time?" *Front. Psychol.* 3:470. doi: 10.3389/fpsyg.2012.00470

Cenni, C., Casarrubea, M., Gunst, N., Vasey, P. L., Pellis, S. M., Wandia, I. N., et al. (2020). Inferring functional patterns of tool use behavior from the temporal structure of object play sequences in a non-human primate species. *Physiol. Behav.* 222:112938. doi: 10.1016/j.physbeh.2021.113498

Clayton, N. S., Bussey, T. J., and Dickinson, A. (2003). Can animals recall the past and plan for the future? *Nat. Rev. Neurosci.* 4, 685–691. doi: 10.1038/nrn1180

Corballis, M. (2019). Language, memory, and mental time travel: an evolutionary perspective. *Front. Hum. Neurosci.* 13:217. doi: 10.3389/fnhum.2019.00217

Dickinson, A. (2012). Associative learning and animal cognition. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 2733–2742.

Dufour, V., and Sterck, E. H. M. (2008). Chimpanzees fail to plan in an exchange task but succeed in a tool using procedure. *Behav. Process.* 79, 19–27. doi: 10.1016/j.beproc.2008.04.003

Feenders, G., and Klump, G. M. (2018). Violation of the unity assumption disrupts temporal ventriloquism effect in starlings. *Front. Psychol.* 9:1386. doi: 10.3389/fpsyg.2018.01386

Hare, B., Call, J., and Tomasello, M. (2001). Do chimpanzees know what conspecifics know? *Anim. Behav.* 61, 139–151. doi: 10.1006/anbe.2000.1518

Healy, S., Bacon, I., Haggis, O., Harris, A., and Kelley, L. (2009). Explanations for variation in cognitive ability: behavioural ecology meets comparative cognition. *Behav. Proc.* 80, 288–294. doi: 10.1016/j.beproc.2008.10.002

Hobaiter, C., Poisot, T., Zuberbuehler, K., Hoppitt, W., and Gruber, T. (2014). Social network analysis shows direct evidence for social transmission of tool use in wild chimpanzees. *PLoS Biol.* 12:9. doi: 10.1371/journal.pbio.1001960

Hoerl, C., and McCormack, T. (2001). *Time and Memory: Issues in Philosophy and Psychology*. Oxford: Oxford University Press.

Hoerl, C., and McCormack, T. (2019). Thinking in and about time: a dual systems perspective on temporal cognition. *Behav. Brain Sci.* 42, 1–77.

Irwin, L. N. (2020). Renewed perspectives on the deep roots and broad distribution of animal consciousness. *Front. Syst. Neurosci.* 14:57. doi: 10.3389/fnsys.2020.00057

Janmaat, K. R. L., Boesch, C., Byrne, R., Chapman, C., Gone Bi, Z., Head, J., et al. (2016). Spatio-temporal complexity of chimpanzee food: how cognitive adaptations can counteract the ephemeral nature of ripe fruit. *Am. J. Primatol.* 78:6. doi: 10.1002/ajp.22527

Janmaat, K. R. L., Polansky, L., Ban, S. D., and Boesch, C. (2014). Wild chimpanzees plan their breakfast time, type and location. *Proc. Natl. Acad. Sci. U.S.A.* 111, 16343–16348. doi: 10.1073/pnas.1407524111

Kaufmann, A. (2015). Animal mental action: planning among chimpanzees. *Rev. Phil. Psych.* 6, 745–760. doi: 10.1007/s13164-014-0228-x

Kaufmann, A. (2016). "Joint distal intentions: who shares what?" in *Routledge Handbook of Philosophy of the Social Mind*, ed. J. Kiverstein (Abingdon: Taylor and Francis), 343–356.

Kaufmann, A., and Cahen, A. (2019). Temporal representation and reasoning in nonhuman animals. *Behav. Brain Sci.* 42:E257. doi: 10.1017/S0140525X19000487

Lameira, A. R., and Call, J. (2018). Time-space-displaced responses in the orangutan vocal system. *Sci. Adv.* 4:11. doi: 10.1126/sciadv.aau3401

Leavens, D., Bard, K., and Hopkins, W. (2010). BIZARRE chimpanzees do not represent "the chimpanzee". *Behav. Brain Sci.* 33, 100–101. doi: 10.1017/s0140525x10000166

Low, P., Panksepp, J., Reiss, D., Edelman, D., Van Swinderen, B., and Koch, C. (2012). "The Cambridge declaration on consciousness," in *Proceedings of the Francis Crick Memorial Conference on Consciousness in Human and non-Human Animals*, (Cambridge: University of Cambridge).

MacLean, E. L., Hare, B., Nunn, C. L., Addessi, E., Amici, F., Anderson, R. C., et al. (2014). The evolution of self-control. *Proc. Natl. Acad. Sci. U.S.A.* 111, e2140–e2148.

Martin-Ordas, G. (2020). It is about time: conceptual and experimental evaluation of the temporal cognitive mechanisms in mental time travel. *Wiley Interdiscip. Rev. Cogn. Sci.* 11:e1530. doi: 10.1002/wcs.1530

Martin-Ordas, G., Haun, D., Colmenares, F., and Call, J. (2020). Keeping track of time: evidence for episodic memory in great apes. *Anim. Cogn.* 13, 331–340. doi: 10.1007/s10071-009-0282-4

Mayo, J. P., and Sommer, M. A. (2013). Neuronal correlates of visual time perception at brief timescales. *Proc. Natl. Acad. Sci. U.S.A.* 110, 1506–1511. doi: 10.1073/pnas.1217177110

Mulcahy, N. J., and Call, J. (2006). Apes save tools for future use. *Science* 312, 1038–1040. doi: 10.1126/science.1125456

Müller, A. J., Massen, J. J. M., Bugnyar, T., and Osvath, M. (2017). Ravens remember the nature of a single reciprocal interaction sequence over 2 days and even after a month. *Anim. Behav.* 128, 69–78. doi: 10.1016/j.anbehav.2017.04.004

Nishida, T., Uehara, S., and Nyondo, R. (1983). Predatory behavior among wild chimpanzees of the Mahale Mountains. *Primates* 20, 1–20. doi: 10.1007/bf02373826

Osvath, M. (2009). Spontaneous planning for future stone throwing by a male chimpanzee. *Curr. Biol.* 19, R190–R191.

Osvath, M., and Karvonen, E. (2012). Spontaneous innovation for future deception in a male chimpanzee. *PloS One* 7:e36782. doi: 10.1371/journal.pone.0036782

Osvath, M., and Martin-Ordas, G. (2014). The future of future oriented cognition in non-humans: theory and the empirical case of the great apes. *Philos. Trans. B Biol. Sci.* 369:20130486. doi: 10.1098/rstb.2013.0486

Osvath, M., and Osvath, H. (2008). Chimpanzee (Pan troglodytes) and orangutan (Pongo abelii) forethought: self-control and pre-experience in the face of future tool use. *Anim. Cogn.* 11, 661–674. doi: 10.1007/s10071-008-0157-0

Osvath, M., Persson, T., and Gärdenfors, P. (2012). Foresight, function representation, and social intelligence in great apes. *Behav. Brain Sci.* 35, 234–235. doi: 10.1017/s0140525x11002068

Pennartz, C. M. A., Farisco, F., and Evers, K. (2019). Indicators and criteria of consciousness in animals and intelligent machines: an inside-out approach. *Front. Syst. Neurosci.* 13:25. doi: 10.3389/fnsys.2019.00025

Perry, C. J., and Chittka, L. (2019). How foresight might support the behavioral flexibility of arthropods. *Curr. Opin. Neurobiol.* 54, 171–177. doi: 10.1016/j.conb.2018.10.014

Redshaw, J., and Suddendorf, T. (2020). Temporal junctures in the mind. *Trends Cogn. Sci.* 24, 52–64. doi: 10.1016/j.tics.2019.10.009

Roberts, W. A., and Feeney, M. C. (2009). The comparative study of mental time travel. *Trends Cogn Sci.* 13, 271–277. doi: 10.1016/j.tics.2009.03.003

Rosati, A. G. (2017). Foraging cognition: reviving the ecological intelligence hypothesis. *Trends Cogn. Sci.* 21, 691–702. doi: 10.1016/j.tics.2017.05.011

Rosati, A. G., Stevens, J. R., Hare, B., and Hauser, M. D. (2007). The evolutionary origins of human patience: temporal preferences in chimpanzees, bonobos, and human adults. *Curr. Biol.* 17, 1663–1668. doi: 10.1016/j.cub.2007.08.033

Safina, C. (2016). Animals think and feel: Précis of beyond words: what animals think and feel (Safina 2015). *Anim. Sentience* 2. doi: 10.51291/2377-7478.1028

Schormans, A. L., Scott, K. E., Vo, A. M. Q., Tyker, A., Typlt, M., Stolzberg, D., et al. (2017). Audiovisual temporal processing and synchrony perception in the rat. *Front. Behav. Neurosci.* 10:246. doi: 10.3389/fnbeh.2016.00246

Smulders, T., Gould, K., and Leaver, L. (2010). Using ecology to guide the study of cognitive and neural mechanisms of different aspects of spatial memory in food-hoarding animals. *Philos. Trans. R. Soc. B.* 365, 883–900. doi: 10.1098/rstb.2009.0211

Spillman, B., van Noordwijk, M. A., Willems, E. P., Mitra Setia, T., Wipfli, U., and van Schaik, C. P. (2015). Validation of an acoustic location system to monitor Bornean orangutan (Pongo pygmaeus wurmbii) long calls. *Am. J. Primatol.* 77, 767–776. doi: 10.1002/ajp.22398

Tomasello, M., and Call, J. (1997). *Primate Cognition*. Oxford: Oxford University Press.

Tulving, E. (2005). "Episodic memory and autonoiesis: uniquely human?," in *The Missing Link in Cognition: Evolution of Self-Knowing Consciousness*, eds H. Terrace and J. Metcalfe (Oxford: Oxford University Press).

van Leeuwen, E. J. C. (2021). Temporal stability of chimpanzee social culture. *Biol. Lett.* 17:20210031. doi: 10.1098/rsbl.2021.0031

van Schaik, C., Damerius, L., and Isler, K. (2013). Wild orangutan males plan and communicate their travel direction one day in advance. *PLoS One* 8:e74896doi. doi: 10.1371/journal.pone.0074896

Viera, G., and Margolis, E. (2019). Animals are not cognitively stuck in time. *Behav. Brain Sci.* 42:e277.

Voelter, C. J., and Call, J. (2014). The cognitive underpinnings of flexible tool use in great apes. *J. Exp. Psychol. Anim. Behav. Process.* 40, 287–302. doi: 10.1037/xan0000025

Webster, M., and Rutz, C. (2020). How STRANGE are your study animals? *Nature* 582, 337–340. doi: 10.1038/d41586-020-01751-5

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Systems Neuroscience

PERSPECTIVE
published: 24 September 2021
doi: 10.3389/fnsys.2021.751876

Consciousness in Jawless Fishes

Daichi G. Suzuki[1,2]*

[1] Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan, [2] Center for Human Nature, Artificial Intelligence, and Neuroscience (CHAIN), Hokkaido University, Sapporo, Japan

Jawless fishes were the first vertebrates to evolve. It is thus important to investigate them to determine whether consciousness was acquired in the common ancestor of all vertebrates. Most jawless fish lineages are extinct, and cyclostomes (lampreys and hagfish) are the sole survivors. Here, I review the empirical knowledge on the neurobiology of cyclostomes with special reference to recently proposed "markers" of primary, minimal consciousness. The adult lamprey appears to meet the neuroanatomical criteria but there is a practical limitation to behavioral examination of its learning ability. In addition, the consciousness-related neuroarchitecture of larvae and its reconstruction during metamorphosis remain largely uninvestigated. Even less is known of hagfish neurobiology. The hagfish forebrain forms the central prosencephalic complex, and the homology of its components to the brain regions of other vertebrates needs to be confirmed using modern techniques. Nevertheless, as behavioral responses to olfactory stimuli in aquariums have been reported, it is easier to investigate the learning ability of the hagfish than that of the lamprey. Based on these facts, I finally discuss the potential future directions of empirical studies for examining the existence of consciousness in jawless fishes.

Keywords: cyclostome, lamprey, ammocoetes, hagfish, minimal consciousness, primary consciousness

OPEN ACCESS

Edited by:
Louis Neal Irwin,
The University of Texas at El Paso, United States

Reviewed by:
Yasunori Murakami,
Ehime University, Japan
Culum Brown,
Macquarie University, Australia

***Correspondence:**
Daichi G. Suzuki
suzuki.daichi.gp@u.tsukuba.ac.jp

Received: *02 August 2021*
Accepted: *02 September 2021*
Published: *24 September 2021*

Citation:
Suzuki DG (2021) Consciousness in Jawless Fishes. Front. Syst. Neurosci. 15:751876. doi: 10.3389/fnsys.2021.751876

INTRODUCTION

The first vertebrates did not have a jaw. These jawless fishes (agnathans) prospered in the Paleozoic, but most of them went extinct (**Figure 1A**). Cyclostomes are the only extant agnathans, consisting of lampreys and hagfish. The jawed vertebrates (gnathostomes) evolved from one of these jawless lineages and then diverged. From a cladistic perspective, the terms "jawless fishes," "jawless vertebrates," and "agnathans" are invalid because they refer to a paraphyletic group. Nevertheless, I use these terms in here for convenience.

Until recently, it was thought that consciousness is limited to the animals with relatively high cognitive ability, such as mammals, birds, and perhaps cephalopods (e.g., Edelman et al., 2005; Edelman and Seth, 2009). However, various researchers have started to consider that all vertebrates, including fishes, share a basic type of consciousness, called primary consciousness or minimal consciousness (Feinberg and Mallatt, 2013, 2016, 2018; Brown, 2015; Bronfman et al., 2016; Ginsburg and Jablonka, 2019; Godfrey-Smith, 2016, 2020). If this is the case, the cyclostomes are important because they are the only remaining stem vertebrates.

Although lampreys and hagfish form a monophyletic group, their brain structures are distinct, reflecting their different lifestyles and lineage-specific adaptations (**Figures 1B–M**). It is thus important to note that modern cyclostomes possess both ancestral and derivative characters. Lampreys spend several years as filter-feeding ammocoetes larvae, which burrow in riverbeds.

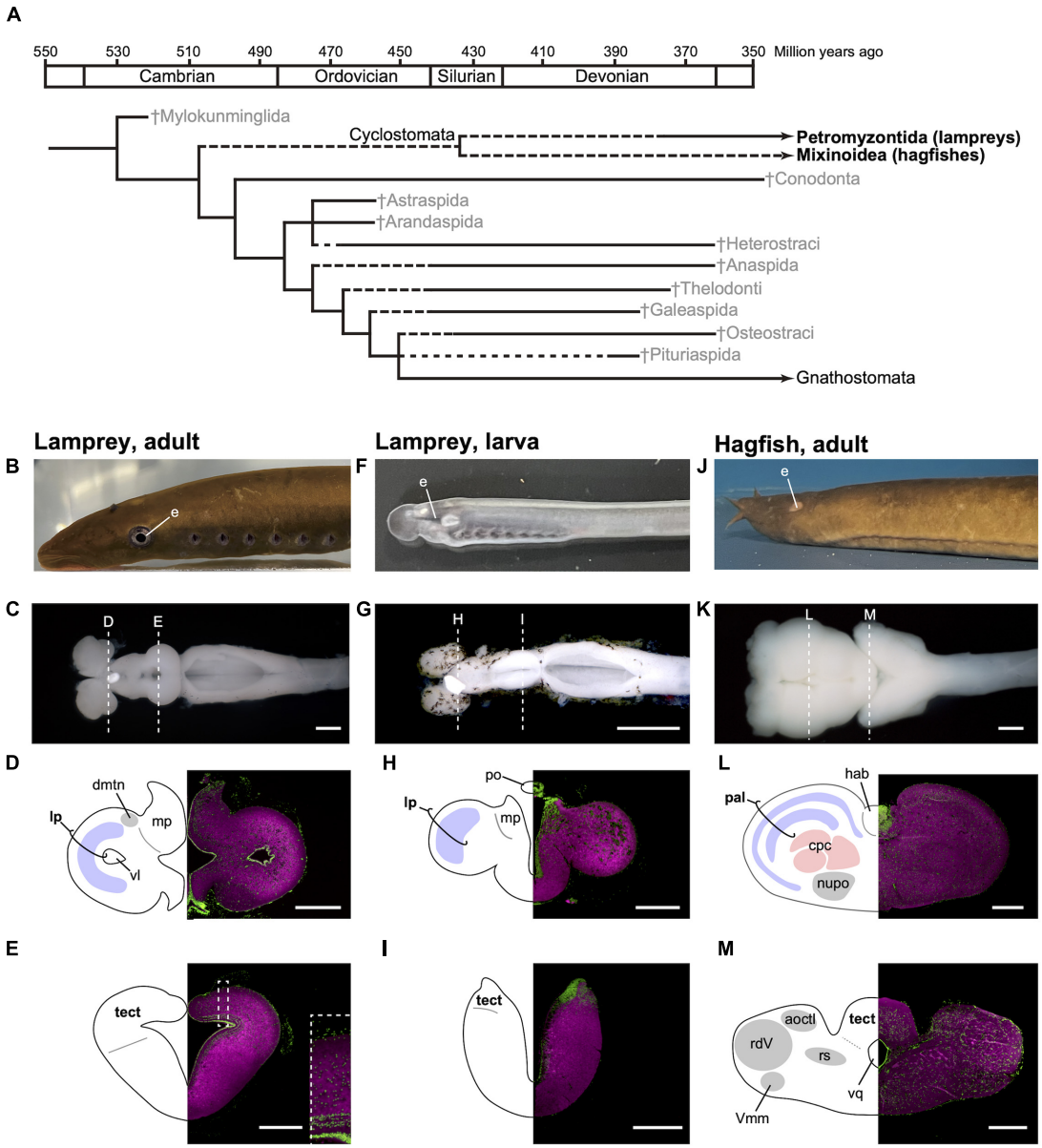


FIGURE 1 | Phylogenetic tree of early vertebrates and brain sections of the cyclostomes. **(A)** Cladogram showing the postulated relationships of the jawless fishes and the Gnathostomata (jawed fishes) based on morphological characters (based on Benton, 2015). **(B–E)** Lateral view of adult lamprey (*Lethenteron camtschaticum*, **B**), dorsal view of the brain **(C)**, and its transverse brain sections at the forebrain **(D)** and midbrain **(E)** levels. The laminated structure of the optic tectum is magnified in the inset of **(E)**. **(F–I)** Lateral view of larval lamprey **(F)**, dorsal view of the brain **(G)**, and its transverse brain sections at the forebrain **(H)** and midbrain **(I)** levels. The photograph for **(G)** is reproduced from Suzuki and Grillner (2018). **(J–M)** Lateral view of adult hagfish (*Eptatretus burgeri*, **J**), dorsal view of the brain **(K)**, and its transverse brain sections at the forebrain **(L)** and midbrain **(M)** levels. Sections are immunostained by anti-acetylated tubulin antibody (Sigma, T6793, *magenta*) and counterstained with Fluorescent Nissl Stain (Invitrogen N21480, *green*). acoctl, area octavolateralis; cpc, central prosencephalic complex; dmtn, dorsomedial telencephalic nucleus; hab, habenular ganglion; lp, lateral pallium; mp, medial pallium; nupo, nuclei praeoptici; pal, pallium; po, pineal organ; tect, tectum; rdV, radix descendens nervi trigemini; rs, formation reticularis, pars superior; vl, ventriculus lateralis; Vmm, nucleus motorius magnocellularis nervi trigemini; vq ventriculus quartus. Scale bars: 1 mm for **(C,G,K)**; 500 μm for **(D,E,L,M)**; 200 μm for **(H,I)**.

As the larva has immature eyes (**Figure 1F**), the optic tectum (the main visual center in non-mammalian vertebrates) also remains undeveloped (**Figure 1I**). On metamorphosis, the animal transforms into an active parasitic predator. Some lampreys are landlocked and breed soon after metamorphosis, while others migrate downstream to the sea or a large lake to attack their prey. The adult lamprey has well-developed eyes (**Figure 1B**) and a mature, layered optic tectum (**Figure 1E**). A recent study found that the lateral pallium of the lamprey has three layers, presumably representing the ancestral vertebrate state,

from which the mammalian cortex is derived (Suryanarayana et al., 2017). In comparison, the hagfish undergoes direct development and has adapted to the deep sea, so its eyes and tectum are degenerate (**Figures 1J,M**). The hagfish forebrain is enlarged (**Figure 1L**) and predominantly receives olfactory input (Wicht and Northcutt, 1993).

In this paper, I explore the current empirical knowledge on the neurobiology of cyclostomes in light of the evolution of consciousness. First, I briefly describe recently proposed "markers" of primary, minimal consciousness. Then, I review current empirical knowledge on the neurobiology of lampreys and hagfish, examining the extent to which the existence of the "markers" is supported in these organisms. Lastly, I discuss possible directions for further studies of consciousness in jawless fishes.

"MARKERS" OF PRIMARY, MINIMAL CONSCIOUSNESS

Among recently proposed accounts of the evolution of consciousness, the theories of Feinberg and Mallatt (2016, 2018), and Ginsburg and Jablonka (2019) are the most detailed and supported by abundant empirical data. In discussing the evolutionary origin of consciousness, the authors use different "markers" of consciousness, while their conclusions are the same; they agree that all vertebrates, as well as some arthropods (including insects) and cephalopods (possibly only coleoids), have consciousness. In this section, I briefly review the two theories and the "markers" of consciousness suggested by these authors.

Feinberg and Mallatt (2016, 2018) distinguish two major aspects of consciousness, exteroceptive and affective consciousness; interoceptive consciousness is intermediate to the two (Feinberg and Mallatt, 2018, Figure 2.4). Their criteria for the exteroceptive consciousness consist of several "special" neurobiological features; complex neural hierarchies (i.e., true brains), isomorphic representations (e.g., somatotopy and retinotopy), multimodal integration ("nested and non-nested hierarchical functions" in their words), interregional neural interactions, and attention. The neuroanatomical and behavioral criteria for affective consciousness include operant learning involving global affective responses and relevant reward/punishment systems [e.g., the ventral tegmental area (VTA) and habenular nucleus].

In contrast to the enumerative approach of Feinberg and Mallatt (2016, 2018), and Ginsburg and Jablonka (2019) argue that a form of associative learning, which they call "unlimited associative learning (UAL)," is *the* positive marker of consciousness. UAL requires a list of capacities (e.g., global accessibility, binding, selective attention, evaluative system, and agency) that suffice for being conscious (Birch et al., 2020). Lacking clear evidence for UAL, they also admit "proxies," including Pavlovian conditioning with compound conditional stimuli, operant conditioning involving novel action patterns, conceptual learning, and navigation learning (Ginsburg and Jablonka, 2019, Table 8.1).

These criteria for consciousness raise two questions. How many of the features listed in the criteria of consciousness proposed by Feinberg and Mallatt (2016) do lampreys and hagfish possess, and do the cyclostomes show UAL or its proxies? In the following sections, I examine these questions applying available empirical evidence.

LAMPREY

The adult lamprey has been used as an experimental model for investigating the basic neuroarchitecture of vertebrates (Grillner et al., 1998; Auclair and Dubuc, 2020), and its neurobiology is relatively well-known. Feinberg and Mallatt (2016) use this knowledge to discuss whether the lamprey has consciousness based on their criteria (pp. 104–115). Current neurobiological findings in fact indicate that the lamprey meets their criteria for exteroceptive consciousness as follows (see also **Table 1**). First, the lamprey brain shares basic brain regions (i.e., the telencephalon, diencephalon, mesencephalon, cerebellum, and rhombencephalon) and developmental mechanisms with other vertebrates (Pombal and Puelles, 1999; Murakami et al., 2001; Pombal et al., 2009; Sugahara et al., 2011, 2016; Murakami, 2017). Second, the optic tectum has a laminar structure, of which the superficial layer receives visual input with retinotopy (Jones et al., 2009). Third, electroceptive inputs are sent to the intermediate layer with spaciotopy, being integrated with visual perception (Kardamakis et al., 2016). In addition, retinotopic and somatotopic organization is found in the lateral portion of the pallium (a telencephalic structure homologous to the mammalian cortex) (Suryanarayana et al., 2020). The lateral pallium sends output to the optic tectum (Ocaña et al., 2015), while the optic tectum sends its fibers to the thalamus (Northcutt and Wicht, 1997), which is the relay center between the pallium/cortex and other brain regions. This suggests that there is a mutual interaction between the pallium and optic tectum (Suzuki and Grillner, 2018, Figure 1C). Lastly, the optic tectum also has mutual connections to the SNc/VTA (SNc: substance nigra pars compacta), which detects the saliency of the visual stimuli and returns the information to the optic tectum via dopaminergic axons (Pérez-Fernández et al., 2017).

Regarding affective consciousness, the lamprey possesses the neuroarchitecture for reward/punishment systems. For example, dopaminergic neurons in the SNc/VTA region send axons not only to the optic tectum (as mentioned above) but also to the basal ganglia, which presumably contributes to reward prediction and motor decision-making based on the prediction (Stephenson-Jones et al., 2011; Pérez-Fernández et al., 2017). The lateral habenula is also present and probably contributes to the reward coding and aversive behavior (Stephenson-Jones et al., 2012; Grillner et al., 2018). The medial habenula sends projections to the interpeduncular nucleus (IPN) and further to the PAG/griseum centrale (PAG: periaqueductal gray) and is perhaps mediates freezing and flight responses (Stephenson-Jones et al., 2012; Grillner et al., 2018). However, little behavioral research has examined learning in the lamprey due to the practical limitation that available adult lampreys

TABLE 1 | The criteria of consciousness and neurobiological evidence in the cyclostomes.

	Lamprey, Adult		Lamprey, Larva		Hagfish	
Feinberg and Mallatt (2016)						
Exteroceptive consciousness						
Complex neural hierarchy (true brain)	*Yes*	Murakami, 2017; Murakami et al., 2001; Sugahara et al., 2011, 2016	*Yes*	Murakami et al., 2001; Murakami, 2017	*Yes*	Larsell, 1947, 1967; Murakami, 2017
		Pombal and Puelles, 1999		Sugahara et al., 2011, 2016		Sugahara et al., 2016, 2017
		Pombal et al., 2009				
Isomorphic representations	*Yes*	Jones et al., 2009; Kardamakis et al., 2016	n.d.	–	*Yes ?*	Amemiya, 1983; Nishizawa et al., 1988
Multimodal integration	*Yes*	Kardamakis et al., 2016	n.d.	–	*Yes ?*	Ronan, 1988; Ronan and Northcutt, 1990; Wicht and Nieuwenhuys, 1998
Interregional neural interaction	*Yes*	Northcutt and Wicht, 1997; Ocaña et al., 2015	n.d.	–	n.d.	–
Attention	*Yes*	Pérez-Fernández et al., 2017	n.d	–	n.d.	–
Affective consciousness						
Operant learning involving global affective response	n.d.	–	n.d.	–	n.d.	–
The relevant reward/punishment system (e.g., VTA, habenular nucleus)	*Yes*	Stephenson-Jones et al., 2011, 2012; Pérez-Fernández et al., 2017; Grillner et al., 2018	n.d.	–	n.d.	–
Ginsburg and Jablonka (2019)						
UAL or its proxies	n.d.	–	n.d.	–	n.d.	–

n.d., not determined.

are postmetamorphic juveniles before downstream migration or mature upstream-migrated fish, both of which lack appetites, making them unsuitable for learning experiments using food rewards. Notably, anadromous adult lamprey can only be alpha conditioned [i.e., conditioning that is based on habituated unconditional stimuli (USs)] and do not show true Pavlovian conditioning when strong lights, strong electric shocks, and nocuous tactile stimulations are used as USs, and weak lights, mild shocks, mild tactile stimuli, sounds, and odors are used and conditional stimuli (CSs) preceding the USs by 3–5 s (Sergeyev, 1964; Razran, 1971).

Interestingly, the lamprey brain changes drastically during postembryonic development. The larval tectum remains immature and becomes laminated during metamorphosis, as mentioned above. The primary retina, which forms during embryogenesis, is also immature and thought to function in non-directional or broadly directional photoreception (Villar-Cerviño et al., 2006; Suzuki et al., 2015a,b; Suzuki and Grillner, 2018). The primary optic nerve projects not to the optic tectum but to the diencephalic pretectum (Suzuki et al., 2015a). A similar neural organization for photoreception is found in amphioxus (Suzuki et al., 2015a), which is a close invertebrate relative of vertebrates and judged to be non-conscious based on the criteria of Feinberg and Mallatt (2016). There are differences in the cytological architecture (discussed in Suzuki et al., 2015a), suggesting a need to analyze the origin of the vertebrate visual system in terms of cell type evolution, possibly with reference to genome duplication in the vertebrate lineage. Nonetheless, the architectural similarity between the two groups implies that the lamprey larval neural circuits for photoreception represent an ancestral state before the evolution of image-forming vision. The marginal region of the primary retina expands into the secondary retina during the entire larval period. The retinal ganglion cells in this secondary retina differentiate before metamorphosis, and the secondary optic nerve projects to the optic tectum with retinotopy (Cornide-Petronio et al., 2011), whereas other retinal cell types (the photoreceptors, horizontal calls, and amacrine cells) differentiate during metamorphosis (De Miguel et al., 1989; Pombal et al., 2003; Villar-Cerviño et al., 2006; Abalo et al., 2008). Thus, the image-forming vision established by the optic tectum is actualized only after the metamorphosis (Suzuki and Grillner, 2018; Suzuki et al., 2019). These findings suggest that the consciousness-related neural circuits are immature during the larval stage and are then reconstructed into the full-blown, functional neuroarchitecture for consciousness during metamorphosis. In other words, the lamprey may undergo transformation from a non-conscious larva to a conscious adult (Suzuki and Grillner, 2018).

Furthermore, the similarity of the neural organization for photoreception between the amphioxus and lamprey larvae implies parallelism between the developmental transformation in the lamprey and the evolutionary transformation in the vertebrate lineage from non-conscious to conscious. However, a recent fossil study indicated that stem lampreys lacked the ammocoetes larval stage (Miyashita et al., 2021), suggesting that the metamorphosis of modern lampreys was acquired secondarily. Evans et al. (2018) agree that ancestral lampreys were direct developers and propose a "condensation hypothesis," which holds that stem lampreys possessed both modern larval and juvenile characters. Differential selection favored segregation of the larval characters in the beginning of the life history and juvenile characters after, requiring metamorphosis to accommodate such body reconstruction. If this is the case, it is possible that stem lampreys gradually developed derivative consciousness-related brain structures, including an image-forming visual system, without evident metamorphosis. Then the development of those structures was condensed in later stages, accompanied by the acquisition of metamorphosis. In either case, the relationship between the evolutionary origin of vertebrate consciousness and the development of lamprey consciousness is an intriguing research topic in terms of evolutionary developmental (evo-devo) biology. Nevertheless, the neural circuits in the larval brain and their transformation during metamorphosis, especially of the optic tectum, remain largely uninvestigated and need further study. The learning ability of the ammocoetes larva is also unknown.

Therefore, the adult lamprey meet the criteria of Feinberg and Mallatt (2016) for exteroceptive consciousness. For affective consciousness, the neuroanatomical criteria are satisfied, although behavioral evidence is lacking. The existence of UAL or its proxies has not been confirmed, thus not meeting the requirement of Ginsburg and Jablonka (2019). The larval lamprey does not appear to satisfy any of the criteria described above, although much more study is needed. If in fact the lamprey changes from non-conscious to conscious during metamorphosis, studies of this transformation will provide valuable information about both the development and evolution of consciousness.

HAGFISH

Much less is known about the neurobiology of the hagfish than that of the lamprey. Although a recent developmental study revealed that the developmental mechanisms underlying formation of the forebrain are conserved in the hagfish (Sugahara et al., 2016), the hagfish forebrain later forms the central prosencephalic complex, and the homology of its components to the brain regions of other vertebrates is unclear (Wicht and Nieuwenhuys, 1998). As a hagfish-specific character, there is no overt epiphysis. A morphologically distinct cerebellum is also absent, while developmental genes involved in cerebellum formation (*Pax6* and *Atoh1*) are expressed in the rhombic lip, from which the cerebellum differentiates (Sugahara et al., 2016, 2017). At the posterior end of the midbrain, there is a portion of the acousticolateral (or vestibulolateral) commissure, which can be regarded as the rudimentary cerebellum (Larsell, 1947, 1967; Sugahara et al., 2017). These findings suggest that the common ancestor of vertebrates possessed at least a non-layered simple cerebellum, similar to that of lampreys.

As mentioned above, the hagfish has degenerate eyes due to adaptation to the deep sea. Fossil evidence indicates that this is a secondary modification specific to the hagfish lineage (Gabbott et al., 2016). In concordance with the degeneration of the eyes, the retinotectal projection is largely reduced, and the retinopretectal pathway becomes dominant (Kusunoki and Amemiya, 1983; Wicht and Northcutt, 1990). Despite no empirical evidence, the degenerate state of the eyes and retinotectal projection implies no or severely disorganized retinotopy in the tectum. Still, it receives inputs from various regions responsible for different sensory modalities (e.g., the octavolateral area, sensory nucleus of the trigeminal nerve, and dorsal column nuclei), suggesting that it functions as an integrative center (Amemiya, 1983; Ronan, 1988; Ronan and Northcutt, 1990; Wicht and Nieuwenhuys, 1998). Furthermore, primary trigeminal afferents are arranged somatotopically in the sensory nucleus of the trigeminal nerve according to the ramus in which they are distributed toward the periphery (Nishizawa et al., 1988). It remains to be determined whether this somatotopic organization is maintained in the tectum. In addition, the hagfish has peculiar taste bud-like chemosensory organs, the Schreiner organs, which are distributed throughout the epidermis and in the prenasal sinus, nasopharyngeal duct, and pharynx at high densities, and in the oral and velar chambers at lower densities (Braun, 1998). These organs are innervated by the trigeminal and glossopharyngeal/vagal nerves and the cutaneous rami of spinal nerves (Braun, 1998). It is plausible that the mechanosensory and chemosensory perception are initially segregated in the primary receptive areas and they are integrated with each other and inputs from other sensory modalities in a higher integrative center. One possibility is that the chemosensory inputs from the Schreiner organs are also received by the tectum. However, these postulates lack solid empirical evidence.

The most prominent sensory modality in the hagfish is olfaction. Its main brain center is the pallium, the forebrain region homologous to the mammalian cortex (Wicht and Northcutt, 1993). The hagfish pallium consists of five layers (Jansen, 1930; Wicht and Northcutt, 1992). Recently, Suryanarayana et al. (2017, 2021) revealed that the lamprey has three layered cortices, which share neuroanatomical and neurophysiological features with those of the reptiles, perhaps being a precursor of the mammalian six-layered neocortex. However, no molecular studies have examined layer-specific genes. Expression analysis on the layer-specific genes is required to elucidate the evolutionary relationships between the five hagfish and three lamprey layers (i.e., which hagfish and lamprey layers correspond), and between the three lamprey layers and the three reptile layers (i.e., whether they are truly homologous or just convergent).

Despite the patchy information, the above findings suggest that the hagfish satisfies some features listed in the criteria of Feinberg and Mallatt (2016) for exteroceptive consciousness

(**Table 1**). However, many of the consciousness-related neuroanatomical features remain to be investigated, including the attention and affective systems.

Still, the hagfish appears to have an advantage in behavioral experiments over the lamprey because it will feed in an aquarium. Recently, Glover et al. (2019) reported that the chemosensory behavior of the hagfish can be assessed using a modified T-maze arena, in which food or noxious stimuli are placed in one of the arms of the maze. This suggests that hagfish learning behavior can be investigated using food as a reward. The degenerate vision of the hagfish is a disadvantage in designing learning experiments. However, odor, taste, and tactile stimuli can be combined to apply compound stimuli, which are required for UAL or its proxies.

CONCLUSION AND FUTURE DIRECTIONS

The cyclostomes are the sole surviving jawless fishes, which were the first vertebrates to evolve. To examine the existence of consciousness in jawless fishes, I assessed knowledge on the neurobiology of the cyclostomes, i.e., lampreys and hagfish, while referring to recently proposed criteria for animal consciousness. The neuroanatomy of the adult lamprey meets the criteria of Feinberg and Mallatt (2016) for exteroceptive consciousness, but much information is lacking.

First, the learning behavior of the adult lamprey needs to be investigated to determine whether the criteria of Feinberg and Mallatt (2016) for affective consciousness are satisfied and whether UAL or its proxies (Ginsburg and Jablonka, 2019) are observed. For this purpose, an innovative experimental design is needed, since available adults do not show appetitive behavior in an aquarium.

Second, the consciousness-related neural circuits in the larval brain and their transformation during metamorphosis, as well as the learning ability of the larva, will be an intriguing subject from the evo-devo perspective on consciousness. Establishment of the multimodal isomorphic (e.g., retinotopic and electroceptive spatiotopic) organization of the optic tectum is of special interest.

Lastly, the neurobiology of the hagfish is less developed in terms of neuroanatomy, neurophysiology, and neuroethology. Further studies using modern approaches, such as gene expression analysis, would improve our understanding of this mysterious creature.

To conclude, we have patchy knowledge on the neurobiology of the cyclostomes for discussing the consciousness of jawless fishes. Despite taxon-specific difficulties in their investigation, further effort is required to elucidate the early evolution of consciousness in the vertebrate lineage.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

DS wrote the manuscript.

FUNDING

This work was supported by the Japan Society for the Promotion of Science (JSPS) under Grants 20K00275 and 20K15855.

REFERENCES

Abalo, X. M., Villar-Cerviño, V., Villar-Cheda, B., Anadón, R., and Rodicio, M. C. (2008). Neurochemical differentiation of horizontal and amacrine cells during transformation of the sea lamprey retina. *J. Chem. Neuroanat.* 35, 225–232. doi: 10.1016/j.jchemneu.2007.12.002

Amemiya, F. (1983). Afferent connections to the tectum mesencephali in the hagfish, *Eptatretus burgeri*: an HRP study. *J. Hirnforsch.* 24, 225–236.

Auclair, F., and Dubuc, R. (2020). "Neural control of swimming in lampreys," in *The Neural Control of Movement*, eds P. J. Whelan and S. A. Sharples (London: Academic Press), 99–123. doi: 10.1016/B978-0-12-816477-8.00005-3

Benton, M. J. (2015). *Vertebrate Palaeontology*, 4th Edn. Oxford: Wiley Blackwell.

Birch, J., Ginsburg, S., and Jablonka, E. (2020). Unlimited Associative Learning and the origins of consciousness: a primer and some predictions. *Biol. Philos.* 35:56. doi: 10.1007/s10539-020-09772-0

Braun, C. B. (1998). Schreiner organs: a new craniate chemosensory modality in hagfishes. *J. Comp. Neurol.* 392, 135–163. doi: 10.1002/(SICI)1096-9861(19980309)392:2<135::AID-CNE1>3.0.CO;2-3

Bronfman, Z. Z., Ginsburg, S., and Jablonka, E. (2016). The transition to minimal consciousness through the evolution of associative learning. *Front. Psychol.* 7:1954. doi: 10.3389/fpsyg.2016.01954

Brown, C. (2015). Fish intelligence, sentience and ethics. *Anim. Cogn.* 18, 1–17. doi: 10.1007/s10071-014-0761-0

Cornide-Petronio, M. E., Barreiro-Iglesias, A., Anadón, R., and Rodicio, M. C. (2011). Retinotopy of visual projections to the optic tectum and pretectum in larval sea lamprey. *Exp. Eye Res.* 92, 274–281. doi: 10.1016/j.exer.2011.01.011

De Miguel, E., Rodicio, M. C., and Anadón, R. (1989). Ganglion cells and retinopetal fibers of the larval lamprey retina: an HRP ultrastructural study. *Neurosci. Lett.* 106, 1–6. doi: 10.1016/0304-3940(89)90192-4

Edelman, D. B., and Seth, A. K. (2009). Animal consciousness: a synthetic approach. *Trends Neurosci.* 32, 476–484. doi: 10.1016/j.tins.2009.05.008

Edelman, D. B., Baars, B. J., and Seth, A. K. (2005). Identifying hallmarks of consciousness in non-mammalian species. *Conscious. Cogn.* 14, 169–187. doi: 10.1016/j.concog.2004.09.001

Evans, T. M., Janvier, P., and Docker, M. F. (2018). The evolution of lamprey (Petromyzontida) life history and the origin of metamorphosis. *Rev. Fish Biol. Fisheries* 28, 825–838. doi: 10.1007/s11160-018-9536-z

Feinberg, T. E., and Mallatt, J. (2013). The evolutionary and genetic origins of consciousness in the cambrian period over 500 million years ago. *Front. Psychol.* 4:667. doi: 10.3389/fpsyg.2013.00667

Feinberg, T. E., and Mallatt, J. M. (2016). *The Ancient Oigins of Consciousness: How the Brain Created Experience*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/10714.001.0001

Feinberg, T. E., and Mallatt, J. M. (2018). *Consciousness Demystified*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/11793.001.0001

Gabbott, S. E., Donoghue, P. C. J., Sansom, R. S., Vinther, J., Dolocan, A., and Purnell, M. A. (2016). Pigmented anatomy in carboniferous cyclostomes and

the evolution of the vertebrate eye. *Proc. R. Soc. Lond. B Biol. Sci.* 283:20161151. doi: 10.1098/rspb.2016.1151

Ginsburg, S., and Jablonka, E. (2019). *The Evolution of the Sensitive Soul: Learning and the Origins of Consciousness*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/11006.001.0001

Glover, C. N., Newton, D., Bajwa, J., Goss, G. G., and Hamilton, T. J. (2019). Behavioural responses of the hagfish Eptatretus stoutii to nutrient and noxious stimuli. *Sci. Rep.* 9:13369. doi: 10.1038/s41598-019-49863-x

Godfrey-Smith, P. (2016). "Animal evolution and the origins of experience," in *How Biology Shapes Philosophy:New Foundations for Naturalism*, ed. D. L. Smith (Cambridge: Cambridge University Press), 51–71. doi: 10.1017/9781107295490.004

Godfrey-Smith, P. (2020). *Metazoa: Animal Minds and the Birth of Consciousness*. London: William Collins.

Grillner, S., Ekeberg, Ö, El Manira, A., Lansner, A., Parker, D., Tegnér, J., et al. (1998). Intrinsic function of a neuronal network: a vertebrate central pattern generator. *Brain Res. Rev.* 26, 184–197. doi: 10.1016/S0165-0173(98)00002-2

Grillner, S., Twickel, A., and Von Robertson, B. (2018). The blueprint of the vertebrate forebrain: with special reference to the habenulae. *Semin. Cell Dev. Biol.* 78, 103–106. doi: 10.1016/j.semcdb.2017.10.023

Jansen, J. A. N. (1930). The brain of *Myxine glutinosa*. *J. Comp. Neurol.* 49, 359–507. doi: 10.1002/cne.900490302

Jones, M. R., Grillner, S., and Robertson, B. (2009). Selective projection patterns from subtypes of retinal ganglion cells to tectum and pretectum: distribution and relation to behavior. *J. Comp. Neurol.* 517, 257–275. doi: 10.1002/cne.22154

Kardamakis, A. A., Pérez-Fernández, J., and Grillner, S. (2016). Spatiotemporal interplay between multisensory excitation and recruited inhibition in the lamprey optic tectum. *Elife* 5:e16472. doi: 10.7554/eLife.16472

Kusunoki, T., and Amemiya, F. (1983). Retinal projections in the hagfish, *Eptatretus burgeri*. *Brain Res.* 262, 295–298. doi: 10.1016/0006-8993(83)91021-1

Larsell, O. (1947). The cerebellum of myxinoids and petromyzonts including developmental stages in the lampreys. *J. Comp. Neurol.* 86, 395–445. doi: 10.1002/cne.900860303

Larsell, O. (1967). *The Comparative Anatomy and Histology of the Cerebellum*. Minneapolis, Min: University of Minnesota Press.

Miyashita, T., Gess, R. W., Tietjen, K., and Coates, M. I. (2021). Non-ammocoete larvae of Palaeozoic stem lampreys. *Nature* 591, 408–412. doi: 10.1038/s41586-021-03305-9

Murakami, Y. (2017). "The origin of vertebrate brain centers," in *Brain Evolution by Design: From Neural Origin to Cognitive Architecture*, eds S. Shigeno, Y. Murakami, and T. Nomura (Japan: Springer), 215–252. doi: 10.1007/978-4-431-56469-0_9

Murakami, Y., Ogasawara, M., Sugahara, F., Hirano, S., Satoh, N., and Kuratani, S. (2001). Identification and expression of the lamprey Pax6 gene: evolutionary origin of the segmented brain of vertebrates. *Development* 128, 3521–3531. doi: 10.1242/dev.128.18.3521

Nishizawa, H., Kishida, R., Kadota, T., and Goris, R. C. (1988). Somatotopic organization of the primary sensory trigeminal neurons in the hagfish, *Eptatretus burgeri*. *J. Comp. Neurol.* 267, 281–295. doi: 10.1002/cne.902670210

Northcutt, R. G., and Wicht, H. (1997). Afferent and efferent connections of the lateral and medial and pallia of the silver lamprey. *Brain Behav. Evol.* 49, 1–19. doi: 10.1159/000112978

Ocaña, F. M., Suryanarayana, S. M., Saitoh, K., Kardamakis, A. A., Capantini, L., Robertson, B., et al. (2015). The lamprey pallium provides a blueprint of the mammalian motor projections from cortex. *Curr. Biol.* 25, 413–423. doi: 10.1016/j.cub.2014.12.013

Pérez-Fernández, J., Kardamakis, A. A., Suzuki, D. G., Robertson, B., and Grillner, S. (2017). Direct dopaminergic projections from the SNc modulate visuomotor transformation in the lamprey tectum. *Neuron* 96, 910–924. doi: 10.1016/j.neuron.2017.09.051

Pombal, M. A., Abalo, X. M., Rodicio, M. C., Anadón, R., and González, A. (2003). Choline acetyltransferase-immunoreactive neurons in the retina of adult and developing lampreys. *Brain Res.* 993, 154–163. doi: 10.1016/j.brainres.2003.09.005

Pombal, M. A., and Puelles, L. (1999). Prosomeric map of the lamprey forebrain based on calretinin immunocytochemistry, nissl stain, and ancillary markers. *J. Comp. Neurol.* 414, 391–422. doi: 10.1002/(SICI)1096-9861(19991122)414:3<391::AID-CNE8>3.0.CO;2-O

Pombal, M. A., Megías, M., Bardet, S. M., and Puelles, L. (2009). New and old thoughts on the segmental organization of the forebrain in lampreys. *Brain Behav. Evol.* 74, 7–19. doi: 10.1159/000229009

Razran, G. (1971). *Mind in Evolution: An East-West Synthesis of Learnt Behavior and Cognition*. Boston, MA: Houghton Mifflin.

Ronan, M. (1988). The sensory trigeminal tract of Pacific hagfish. *Brain Behav. Evol.* 32, 169–180. doi: 10.1159/000116544

Ronan, M., and Northcutt, R. G. (1990). Projections ascending from the spinal cord to the brain in petromyzontid and myxinoid agnathans. *J. Comp. Neurol.* 291, 491–508. doi: 10.1002/cne.902910402

Sergeyev, B. F. (1964). The structure of temporary connections in lower chordates. *Zhurnal Vysshey Nervnoy Deyatel'nosti imeni I. P. Pavlova* 14, 904–910.

Stephenson-Jones, M., Floros, O., Robertson, B., and Grillner, S. (2012). Evolutionary conservation of the habenular nuclei and their circuitry controlling the dopamine and 5-hydroxytryptophan (5-HT) systems. *Proc. Natl. Acad. Sci. U. S. A.* 109, E164–E173. doi: 10.1073/pnas.1119348109

Stephenson-Jones, M., Samuelsson, E., Ericsson, J., Robertson, B., and Grillner, S. (2011). Evolutionary conservation of the Basal Ganglia as a common vertebrate mechanism for action selection. *Curr. Biol.* 21, 1081–1091. doi: 10.1016/j.cub.2011.05.001

Sugahara, F., Aota, S., Kuraku, S., Murakami, Y., Takio-Ogawa, Y., Hirano, S., et al. (2011). Involvement of Hedgehog and FGF signalling in the lamprey telencephalon: evolution of regionalization and dorsoventral patterning of the vertebrate forebrain. *Development* 138, 1217–1226. doi: 10.1242/dev.059360

Sugahara, F., Murakami, Y., Pascual-Anaya, J., and Kuratani, S. (2017). Reconstructing the ancestral vertebrate brain. *Dev. Growth Differ.* 59, 163–174. doi: 10.1111/dgd.12347

Sugahara, F., Pascual-Anaya, J., Oisi, Y., Kuraku, S., Aota, S. I., Adachi, N., et al. (2016). Evidence from cyclostomes for complex regionalization of the ancestral vertebrate brain. *Nature* 531, 97–100. doi: 10.1038/nature16518

Suryanarayana, S. M., Pérez-fernández, J., Robertson, B., and Grillner, S. (2021). The lamprey forebrain: evolutionary implications. *Brain Behav. Evol.* 1–16. doi: 10.1159/000517492 [Epub ahead of print].

Suryanarayana, S. M., Pérez-Fernández, J., Robertson, B., and Grillner, S. (2020). The evolutionary origin of visual and somatosensory representation in the vertebrate pallium. *Nat. Ecol. Evol.* 4, 639–651. doi: 10.1038/s41559-020-1137-2

Suryanarayana, S. M., Robertson, B., Wallén, P., and Grillner, S. (2017). The lamprey pallium provides a blueprint of the mammalian layered cortex. *Curr. Biol.* 27, 3264.e–3277.e. doi: 10.1016/j.cub.2017.09.034

Suzuki, D. G., and Grillner, S. (2018). The stepwise development of the lamprey visual system and its evolutionary implications. *Biol. Rev.* 93, 1461–1477. doi: 10.1111/brv.12403

Suzuki, D. G., Murakami, Y., Escriva, H., and Wada, H. (2015a). A comparative examination of neural circuit and brain patterning between the lamprey and amphioxus reveals the evolutionary origin of the vertebrate visual center. *J. Comp. Neurol.* 523, 251–261. doi: 10.1002/cne.23679

Suzuki, D. G., Murakami, Y., Yamazaki, Y., and Wada, H. (2015b). Expression patterns of *Eph* genes in the "dual visual development" of the lamprey and their significance in the evolution of vision in vertebrates. *Evol. Dev.* 17, 139–147. doi: 10.1111/ede.12119

Suzuki, D. G., Pérez-Fernández, J., Wibble, T., Kardamakis, A. A., and Grillner, S. (2019). The role of the optic tectum for visually evoked orienting and evasive movements. *Proc. Natl. Acad. Sci. U.S.A.* 116, 15272–15281. doi: 10.1073/pnas.1907962116

Villar-Cerviño, V., Abalo, X. M., Melendez-ferro, M., Perez-Costas, E., Holstein, G. R., Rodicio, M. C., et al. (2006). Presence of glutamate, glycine, and γ- aminobutyric acid in the retina of the larval sea lamprey: comparative Immunohistochemical study of classical Neurotransmitters in larval and postmetamorphic retinas. *Comp. Gen. Pharmacol.* 499, 810–827. doi: 10.1002/cne.21136

Wicht, H., and Nieuwenhuys, R. (1998). "Hagfishes (Myxinoidea)," in *The Central Nervous System of Vertebrates*, Vol. 1, eds R. Nieuwenhuys, H. J. Ten Donkelaar, and C. Nicholson (Berlin: Springer Berlin Heidelberg), 497–549. doi: 10.1007/978-3-642-18262-4_11

Wicht, H., and Northcutt, R. G. (1990). Retinofugal and retinopetal projections in the Pacific hagfish, Eptatretus stouti (Myxinoidea). *Brain Behav. Evol.* 36, 315–328. doi: 10.1159/000115317

Wicht, H., and Northcutt, R. G. (1992). The forebrain of the Pacific hagfish: a cladistic reconstruction of the ancestral craniate forebrain. *Brain Behav. Evol.* 40, 25–64. doi: 10.1159/000108540

Wicht, H., and Northcutt, R. G. (1993). Secondary olfactory projections and pallial topography in the pacific hagfish, *Eptatretus stouti*. *J. Comp. Neurol.* 337, 529–542. doi: 10.1002/cne.903370402

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 24 September 2021
doi: 10.3389/fpsyg.2021.732336

Multiple Routes to Animal Consciousness: Constrained Multiple Realizability Rather Than Modest Identity Theory

Jon Mallatt[1]* *and Todd E. Feinberg*[2]

[1]*The University of Washington WWAMI Medical Education Program at The University of Idaho, Moscow, ID, United States,*
[2]*Department of Psychiatry and Neurology, Icahn School of Medicine at Mount Sinai, New York, NY, United States*

The multiple realizability thesis (MRT) is an important philosophical and psychological concept. It says any mental state can be constructed by multiple realizability (MR), meaning in many distinct ways from different physical parts. The goal of our study is to find if the MRT applies to the mental state of consciousness among animals. Many things have been written about MRT but the ones most applicable to animal consciousness are by Shapiro in a 2004 book called The Mind Incarnate and by Polger and Shapiro in their 2016 work, The Multiple Realization Book. Standard, classical MRT has been around since 1967 and it says that a mental state can have *very many* different physical realizations, in a nearly unlimited manner. To the contrary, Shapiro's book reasoned that physical, physiological, and historical constraints force mental traits to evolve in just a few, limited directions, which is seen as convergent evolution of the associated neural traits in different animal lineages. This is his mental constraint thesis (MCT). We examined the evolution of consciousness in animals and found that it arose independently in just three animal clades—vertebrates, arthropods, and cephalopod mollusks—all of which share many consciousness-associated traits: elaborate sensory organs and brains, high capacity for memory, directed mobility, etc. These three constrained, convergently evolved routes to consciousness fit Shapiro's original MCT. More recently, Polger and Shapiro's book presented much the same thesis but changed its name from MCT to a "modest identity thesis." Furthermore, they argued against almost all the classically offered instances of MR in animal evolution, especially against the evidence of neural plasticity and the differently expanded cerebrums of mammals and birds. In contrast, we argue that some of these classical examples of MR are indeed valid and that Shapiro's original MCT correction of MRT is the better account of the evolution of consciousness in animal clades. And we still agree that constraints and convergence refute the standard, nearly unconstrained, MRT.

Keywords: animal consciousness, multiple realizability, convergent evolution, mental constraint thesis, modest identity thesis, compensatory differences, mental phenomena, evolutionary constraints

OPEN ACCESS

Edited by:
Peter Beim Graben,
Humboldt University of Berlin, Germany

Reviewed by:
George F. R. Ellis,
University of Cape Town, South Africa
Kurt Kotrschal,
University of Vienna, Austria

***Correspondence:**
Jon Mallatt
jmallatt@uidaho.edu

Specialty section:
This article was submitted to Consciousness Research, a section of the journal Frontiers in Psychology

Received: *28 June 2021*
Accepted: *13 August 2021*
Published: *24 September 2021*

Citation:
Mallatt J and Feinberg TE (2021) Multiple Routes to Animal Consciousness: Constrained Multiple Realizability Rather Than Modest Identity Theory. Front. Psychol. 12:732336. doi: 10.3389/fpsyg.2021.732336

INTRODUCTION

Our research program focuses on which animals have at least a minimal or primary form of consciousness; that is, have raw, nonreflective experiences of images constructed from sensing the world and also experience affects, meaning emotions, and moods (Mallatt and Feinberg, 2020; Mallatt et al., 2021). We have worked together on this program for almost a decade (Feinberg and Mallatt, 2013, 2016, 2018, 2019, 2020). In our work, we use systems theory to argue that consciousness is an evolved product of complex brains in complex bodies, so it is an emergent feature of a complex physical system (Feinberg and Mallatt, 2020). One feature of every emergent, complex system is that its end-process can be caused in multiple ways or by "multiple routes" (reviewed by Feinberg and Mallatt, 2020). Examples of this multiple-routes feature are: waves emerging in a body of water, which can be caused by either the wind, a stone, or an earthquake; a traffic jam, which can be caused by bad weather, too many vehicles on the road, or an accident ahead; and the patterns formed by "cellular automata," which are computer simulations programed to follow various rules (Bedau, 2008, p. 180).

In the study of the mind, this multiple-routes feature has been called *multiple realizability (MR),* as promoted by the *multiple realizability thesis (MRT)*, which says that a mind and its mental states can be constructed in many distinct ways from different physical parts (Bickle, 2020).[1,2] MRT is of philosophical importance for addressing the mind-body problem because it is at the core of the dominant philosophical view called non-reductive physicalism, which says that mental states have strictly physical causes but do not reduce to counterparts in the more basic sciences, such as physics and neurobiology (Kim, 2008; Macdonald and Macdonald, 2019). Here, we emphasize the *multiple* in multiple realizability. Indeed, as Bickle (2020) points out, the most popular versions of MRT, named the "standard" and "radical" versions, say that *very many* types of physical states can cause, or realize, the same mental state.

MRT was constructed (Putnam, 1967) to refute the mind-brain *identity theory*, which says that all mental states are identical to brain states, and which itself arose as a solution to the mystery of how the mind relates to the brain (Place, 1956; Smart, 1959). The logic by which MRT argues against the identity theory is that if a mental state has many different causes; then, it has no single cause so we cannot look for any identity or even generality among the causes of the state. Each instance could have a different cause, with the causes having no physical properties in common (Baysan, 2019; Bickle, 2020).

MRT asserts—and we agree—that mental phenomena or states really do exist as mental kinds. That is, in accordance with the disciplines of psychology and neuroscience, MRT recognizes such general kinds as explicit memory, feeling acute pain, associative learning, cognitive problem solving, and consciousness, with each kind occurring as the same thing in different humans and different species. Calling these things "kinds" can always be opposed, philosophically, by successive "kind splitting" (Aizawa, 2013; Polger and Shapiro, 2016, pp. 99–104), where the opponent argues that the claimed mental state (sharp pain or memory, for example) is not the same in a rat as in a human, in a monkey as in a human, in two different humans, or in the same human at two different times. As evolutionary biologists, we resist such kind-splitting on the grounds that the mental kinds have adaptive value in multiple taxa of brainy animals. We reason that strong selection pressures demand the psychological states be the same for the different taxa to survive in competition in the same, real world. Any competing taxon without memory or attention skills would quickly go extinct.

We definitely include consciousness among the mental kinds that are shared by different taxa (Ben-Haim et al., 2021). The evidence for this that impressed us most was from Neider et al. (2020), in which crows demonstrated human-derived markers of consciousness (single-neuron responses that mark visual perception) at the same time these crows showed monkey-like cognitive skills (the ability to report their perceptions). This was good evidence for the conscious mental kind across the distantly related birds and mammals.

The present paper focuses on the studies of Lawrence Shapiro and Thomas Polger because they are the authors in the MR field who most closely considered the mental states of animals—animal consciousness being the theme of this special issue. Shapiro chose not to include consciousness among the states he analyzed for multiple realizability because consciousness has difficult, subjective aspects (Shapiro, 2004, pp. 70, 228–229). However, we see consciousness as a valid mental state that has been defined well enough and can be studied analytically and scientifically (Nagel, 1974; Mallatt and Feinberg, 2020; Mallatt et al., 2021; Mallatt, 2021a). Therefore, we will go ahead and analyze whether it is a multiply-realized phenomenon in the animal kingdom. That is the goal of this paper.

PART 1: SHAPIRO ON MULTIPLE REALIZATION AND CONSTRAINTS

The Importance of Evolutionary Constraints

Because MRT claims that so many different physical mechanisms can give rise to each mental state, Shapiro investigated whether this claim fits biological reality. If, as MRT asserts, the same state has little in common across the animal taxa in its causal

[1]The related term, multiple *realization*, is also used. "Multiple realizability" refers to all the possible physical causes (including imaginary ones), whereas "multiple realization" refers only to the known physical candidates like neurons and brains (Bickle, 2020). We will not make this realizability-realization distinction, however, because the literature we are reviewing seldom distinguishes the terms.
[2]Although "multiple realizability" was originally applied only to mental states, now this term is also being applied to the functional states of other complex physical systems (Ellis, 2012). These multiply-realized states range from convection currents in liquids (Bishop and Silberstein, 2019), to protein biochemical states (Tahko, 2020), to the elasticity of different polymers (McLeish, 2019), to transitions at a critical point in fluids and ferromagnets (Blundell, 2019), and to electrical wires and more (Aizawa, 2013).

mechanisms, then "we should be able to make few predictions about the properties of the organ that realizes" a mental state (Shapiro, 2004, p. 137). Next, Shapiro continues, MRT claims that the functions of any state place few constraints on the properties that can cause such a state. With so few constraints, therefore, MRT also predicts there will be no or little convergent evolution of the structures related to any mental kind across distantly related taxa. To the contrary, convergent evolution is common (Conway Morris, 2003; McGhee, 2019), and Shapiro refuted these MRT predictions by documenting many examples of it, as channeled by physical, physiological, and historical constraints (also see Vogel, 1998). Shapiro's best examples are convergently evolved similarities in different eyes and the independent evolution of modular subparts in the brains of different animals. (We document these below.)

The many documented instances where constraints produced convergent evolution led Shapiro to reject MRT as wrong for claiming that "almost anything goes." He replaced MRT with his mental constraint thesis (MCT). This thesis says a given mental state can have only a few types of neural causes (realizers)—far fewer than allowed by standard MRT, a "handful" rather than "hundreds or thousands" (p. 32). He illustrated MCT with helpful analogies. Mechanical devices for removing the cork from a wine bottle (pp. 1–2, 46–51, 68) are constrained to those that pull, suck, push, or twist out the cork, because not much else will work. A bit that drills through rocks for oil can only consist of diamond or hardened metal and it invariably uses a rotatory action. Without these constraints, the bit could not penetrate the rock fast enough or would wear out too soon. Only two types of bits fit the necessary conditions: the rolling cutter bit and the fixed cutter bit (**Figure 1**).

Does the mental state of primary consciousness fit MRT or does it fit MCT? To answer this, we must provide some background. In our prior studies, we deduced that only three clades of animals are conscious (Feinberg and Mallatt, 2016, 2018; Mallatt, 2021a,b). This deduction came from two reasoned assumptions: (1) an animal has consciousness if it builds detailed, multisensory representations of the world with mapped, topographically arranged neural pathways to and in its brain and (2) if it is capable of elaborate operant learning from rewards and punishments.[3] The only animals that fit these criteria are all the vertebrates, all the arthropods, and cephalopod mollusks (octopus, squid, and cuttlefish). Importantly, these unrelated taxa share many consciousness-related features, which are listed in **Table 1**. The small number of conscious taxa—just three—indicates evolutionary convergence with constraints, MCT not MRT. We emphasize that the vertebrates, arthropods, and cephalopods fully fit the criteria for convergently evolved consciousnesses, having descended independently from a distant common ancestor that lacked a brain and was without consciousness (Northcutt, 2012; Feinberg and Mallatt, 2016; **Figure 2**).

The fact that the MRT did not consider the constraints or the limitations that these constraints impose seems a major blind spot of that thesis. As Shapiro (2004, p. 21–23) pointed out, constraints are faced by every living system that has goal-directed functions because unless such a system is constructed in a certain, constrained way it cannot perform its function. Consciousness certainly meets this criterion of having an adaptive function that benefits survival (Cabanac, 1996; Seth, 2009; Feinberg and Mallatt, 2018), its function being to aid decision making by allowing one to consider alternate choices. Stated another way, consciousness processes complex sensory information to choose and direct the movements of large, multicellular bodies in space, for finding food and mates and for escaping danger (**Table 2**). The constraints necessary for this function are needing neurons, sensory organs, muscles, and many more.

This is not to deny that many differences exist among the nervous systems of vertebrates, arthropods, and cephalopods, along with their similarities. Their brains look different and the analogous functional areas do not have the same relative locations in the brains (**Figure 3**). The similarities still abound, however, so constraints have channeled the emergence of these conscious systems into similar directions (**Table 1**).

Because the shared features in **Table 1** provide real empirical support for Shapiro's MCT, we will examine several of them to show how strong the constraints can be for the evolutionary convergence of conscious systems. These constrained features are sensory systems, brain organization, mapping, valence neurons, and memory systems. Note as we present them that these features are not unique to conscious systems and animals, but they are necessary for consciousness, and they are much better developed in the conscious animals than in nonconscious animals. So are the *functions* of these features. Thus, they will be informative about consciousness and its constraints.

Mental Constraints on Conscious Systems

Constraints on Sensory Systems

For consciousness to play its role of sensing and mapping the environment in detail, it must have sensory receptors and sensory pathways for all the classes of stimuli: light, mechanical forces, smells, tastes, and temperature. To operate efficiently, these structures must register the location and intensity of each stimulus, and they enhance the contrast between nearby stimuli by a process called lateral inhibition (Shapiro, 2004, Chapter 4). These properties can be seen as constraints that led to convergent evolution because they characterize *all* vertebrates, arthropods, and cephalopods (Hartline et al., 1956; Nahmad-Rohen and Vorobyev, 2019; Kandel et al., 2021).

Similarities in the image-forming *eyes* of the three taxa are especially noteworthy. Vertebrates and cephalopods have "simple," spherical camera eyes that are the most alike, remarkably so considering they evolved independently (**Figure 4**). Many of the similarities in the lenses, dimensions, and compartments of these two eyes serve to eliminate spherical aberration, a lens problem that blurs the image (Shapiro, 2004, pp. 99–104).

[3]Our reasoning behind these two assumptions is: (1) it is logical to say that if an animal exerts the energy to build the detailed and mapped representations, then it will use these representations as mental reference images to help it move and operate in the real world; (2) complex operant learning seems to provide two-part evidence that an animal feels emotions, both the initial attraction to a reward and then recalling the reward-feeling that motivates the learned behavior (Mallatt et al., 2021).

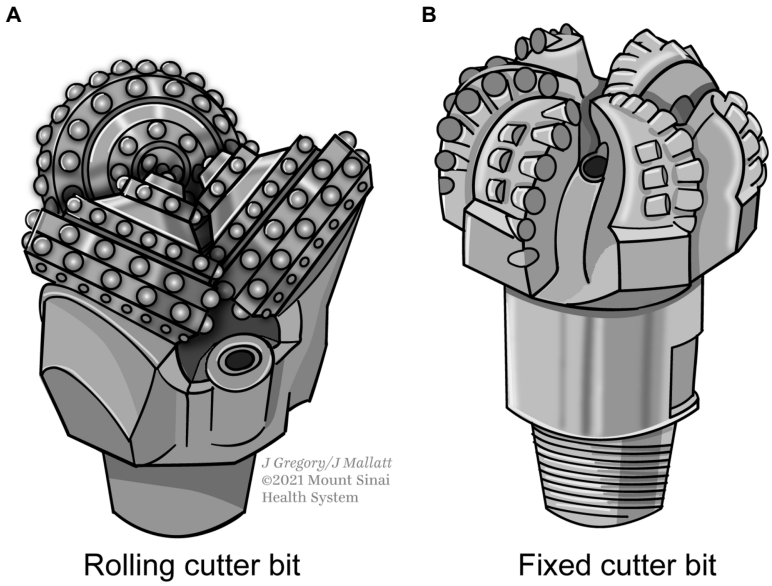


FIGURE 1 | Constraints in the design of rock-drilling bits for the petroleum industry. Only two types **(A)** and **(B)** are practical for this purpose. For photos, see https://petgeo.weebly.com/types-of-drilling-bits.html.

Along with the similarities, the eyes of the conscious clades show some differences. For example, arthropod eyes are not simple but compound, made of many tube-like *ommatidia*. They differ from camera eyes in some significant ways, the retina being convex instead of concave and in having many lenses instead of one. This design sacrifices some visual acuity but is better for detecting movement and it gives the eye a wider field of view.

Despite these differences, compound and simple eyes share many similarities that are demanded by the constraints for image formation: corneas, lenses, and photoreceptor cells. Additionally, the visual pathway from the eye photoreceptors to the visual brain centers is remarkably similar in arthropods and vertebrates (Sanes and Zipursky, 2010). For cephalopods, the visual pathway is far less studied, but it resembles that of arthropods and vertebrates in having especially many levels of successive neurons (Feinberg and Mallatt, 2016; Table 9.2). To summarize this topic, the many similarities between the eyes and visual pathways of the three taxa indicate that only a limited number of structures can produce formed images, favoring Shapiro's MCT over MRT.

Constraints on Brain Organization

A central nervous system contains information-processing neural networks whose neurons are connected by "wires" or "cables" in the form of neuronal axons and dendrites. In such a system, it costs energy to connect and use the wires, so natural selection acts to minimize the cost, especially by minimizing the total length of the wires (Shapiro, 2004, pp. 124–132). Computer simulations show that the best way to minimize cost and maximize fitness is to partition the many neurons into modular groups (local neuronal processing centers), with each *module* having many internal connections but fewer connections to other modules (Simon, 2002). In this way, each module can perform its special processing function and then send a condensed summary out to other parts of the network. A *hierarchical* organization will also emerge, in which the modules have submodules so that each submodule solves a *part* of the module's processing task (Mengistu et al., 2016). Modularity makes brains more evolutionarily adaptable because "swapping or rearranging maladaptive modules is less costly than rearranging the entire system" (Sporns and Betzel, 2016). Furthermore, having a hierarchy of modules helps to keep a neuronal system in a balanced "critical state," where the local electrical activity can persist, neither dying out nor spreading uncontrollably through the whole system (Kaiser et al., 2007; Rubinov et al., 2011).

This ideal, modular arrangement takes its highest form in the brains of conscious animals, matching the arrangement we deduced for consciousness (Feinberg and Mallatt, 2019, 2020). We described it as a hierarchical organization with many neural computing modules and networks that are distributed but integrated, for both local functional specialization and coherence among the many parts of the brain. Given this match, these must be constraints that directed the evolutionary emergence of a conscious brain, as happened in vertebrates, arthropods, and cephalopods. Several sources document these traits of increased hierarchy, modularity (brain nuclei and laminae), and fiber connections (tracts and neuropils) in all three of the clades: in the vertebrates (Striedter and Northcutt, 2020), arthropods (plus their nearest relatives the velvet worms: Strausfeld, 2012), and cephalopods (Shigeno et al., 2018; Wang and Ragsdale, 2019). Once more we have uncovered multiple constraints that led to convergent evolution, as Shapiro's MCT predicted.

TABLE 1 | The convergently evolved features of consciousness that are shared by vertebrates, arthropods, and cephalopod mollusks (mostly from Feinberg and Mallatt, 2020).

Neural complexity (more than in a simple, core brain)
• Brain with many neurons (>100,000?)
• Many subtypes of neurons
Elaborated sensory organs
• Image-forming eyes, receptors for touch, taste, hearing, smell
Neural hierarchies with neuron–neuron interactions
• Extensive reciprocal communication in and between pathways for the different senses
• Brain's neural computing modules and networks are distributed but integrated, leading to local functional isolation plus global coherence
• Synchronized communication by brain-wave oscillations
• Neural spike trains form representational codes
• The higher brain levels allow the complex processing and unity of consciousness
• Higher brain levels exert considerable influence on the lower levels such as motor neurons, for top-down causality
• Hierarchies that let consciousness predict events a fraction of a second in advance
Pathways that create mapped mental images or affects (affects being emotions and moods)
• Neurons are arranged in topographic maps of the outside world and body structures
• Valence coding of good and bad, for affective states
• Feed into pre-motor brain regions to motivate, choose, and guide movements in space for high mobility
Brain mechanisms for selective attention and arousal
Memory of perceived objects or events

TABLE 2 | Some adaptive roles of consciousness (from Feinberg and Mallatt, 2019) that constrain the types of features that can produce this phenomenon.

• Consciousness organizes large amounts of sensory input into a set of phenomenal properties for choosing which actions to perform
• Its unified simulation of the sensed world directs behavior in this world
• It ranks sensed stimuli by importance, by assigning affects to them, making decisions easier (Cabanac, 1996)
• Allows flexible behavior because it sets up many different behavioral choices
• Allows easily adjustable behavior because it predicts the consequences of one's actions into the immediate future (Perry and Chittka, 2019; Solms, 2019)
• Deals well with new situations, to meet the changing challenges of complex environments

Constraint of Valence Neurons and Circuits

Having some way to encode value, "good" and "bad," is a necessity for the affective (~emotional) feelings of consciousness. The existence of value (valence) neurons and circuits is well documented in the brains of vertebrates (Berridge and Kringelbach, 2015; Betley et al., 2015; Namburi et al., 2016; Panksepp, 2016; Tye, 2018), and valence circuits also have been found in arthropods (Felsenberg et al., 2017; Eschbach et al., 2020a,b; Siju et al., 2020). They have not been sought in cephalopods. We should make clear that we are not claiming valence neurons and circuits *explain* good and bad feelings. They are just part of the realizer mechanism that leads to such consciousness.

Constraint of Memory Systems

A conscious animal requires a good deal of memory in order to navigate through space using recalled landmarks and in order to learn extensively from past experiences. For these functions of consciousness, memory storage would have to exist in the form of mental representations about the features of this world relevant for a certain species or individual, organized in a more or less episodic way. Thus, we reason that all conscious animals must have relatively large brain regions for memory. This prediction proves true (**Figure 3**). Vertebrate brains have large memory regions, such as the hippocampus and amygdala (Brodal, 2016), arthropod brains have mushroom bodies for memory (Strausfeld, 2012), and in cephalopod brains, the frontal and vertical lobes participate in sensory memory (Shigeno et al., 2018; Figure 3 in Wang and Ragsdale, 2019). In the three clades, the functional constraints of consciousness independently directed their brains to evolve toward increased memory storage. Once more, constraints led to convergent evolution, as MCT predicts.

Conclusion of Part 1

Shapiro's (2004, p. 137–138) book asked whether future empirical research will show if his mental constraint thesis is more valid than the largely unconstrained MRT. Our findings on the convergent evolution of consciousness in vertebrates, arthropods, and cephalopods provide an answer, indicating that MCT is indeed more valid. We accept MCT as better than MRT not only because it fits our own findings but also because unlike standard MRT it incorporates convergent evolution, an important part of evolutionary theory.

Since 2004 Shapiro has developed more ideas on multiple realizability (Shapiro, 2008; Shapiro and Polger, 2012), and he coauthored a book on this subject as Polger and Shapiro (2016). Thus, we must examine that book to see whether these authors' ideas on MCT have changed and if we still favor them.

PART 2: POLGER AND SHAPIRO ON MULTIPLE REALIZATION: IDENTITY THEORY AFTER ALL?

Points of Agreement

A theme of Polger and Shapiro's (2016) Multiple Realization Book, henceforth called "P and S," is that the best explanation of mental processes makes some use of mind-brain identities in a "*modest identity theory*," meaning that instances of multiple realization are less common than many philosophers assume (pp. 34, 144–145). This turned out to be a logical and direct extension of the authors' previous ideas on MR. Close reading shows P and S did come to the same conclusion as Shapiro (2004), the conclusion that constraints led to the same mental kinds evolving convergently, with similar neural realizers, in just a few different taxa and that the constraints refute the standard, unconstrained multiple realization thesis (P and S, p. 143).

However, P and S went a step beyond the earlier MCT by calling their new version an identity theory, although one that

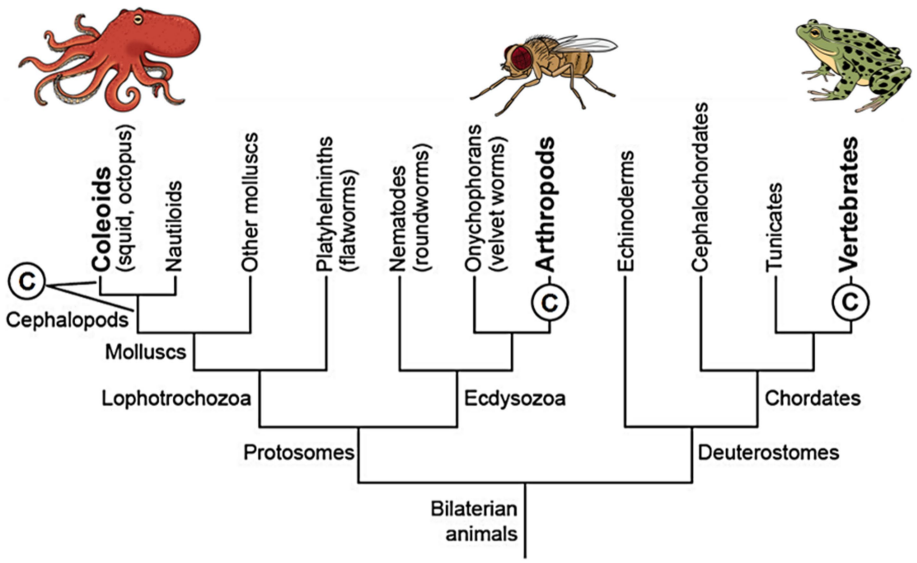


FIGURE 2 | A simplified phylogenetic tree of animal relationships showing that consciousness (©) emerged independently in three different lines of animals. At left, the two leaders extending from the © mean that we cannot tell whether the consciousness evolved in the first cephalopod mollusks or else in the coleoid ancestor of squid, octopus, and cuttlefish. Reproduced with the permission of the copyright holder Mount Sinai Health System.

still allows some mental kinds to be multiply realized (pp. 144–145). Their reason for calling it an identity theory seems to be as follows (p. 143). All brains are complex and more complexity imposes more constraints on the types of neuromechanisms that can perform a given mental function; thus, the complex psychological functions must be realized in "very similar ways" in differently evolved brains due to all the constraints. So far, we can follow their logic, but then P and S apparently equated "very similar ways" with "*identical* ways" to reach their identity theory. That is, they concluded that similarly constrained, convergent solutions are effectively identical solutions. To the contrary, we view "very similar" solutions as nonidentical so we do not consider this—nor the original MCT idea—to be an identity theory. Rather, we see these solutions as highly constrained versions of the multiple realization thesis. Our disagreement, however, may be merely semantic hair-splitting because both we and P and S agree that our two interpretations fall on a spectrum and are close together on this spectrum. That is, there may be no practical difference between our "highly constrained MRT" and their "modest identity theory that allows some MR."

This means that we and P and S would be in agreement—except for one more thing. They devoted much of their book to arguing against almost every case that has ever been used to support MRT. By contrast, we judge that many of these cases validly support MRT (albeit the constrained version of multiple realizability to which we subscribe).

Points of Disagreement

The anti-MR cases in question involve (1) neural plasticity, (2) ideas about compensatory differences in mental kinds, and (3) comparing the brains of birds and mammals. Before looking at these cases, however, we must point out that P and S developed valuable and rigorous criteria for judging whether a test case truly indicates MR—an undertaking that has always been difficult and confusing. Here in paraphrased form are their criteria, which they called their Official Recipe (P and S, p. 67):

1. The realized mental kind must be the same in the animals being compared.
2. The realizers must be different.
3. The differences between the realizers must make the kind the same in the two animals.
4. The differences between the realizers cannot be trivial: They cannot be merely the differences one sees within a mental kind.

Although this Official Recipe nicely formalizes the decision process and helps to refute some cases that were wrongly said to support MRT, it cannot always provide certainty. Judgment calls will still remain over whether the kinds are really the same in two individuals (in criterion 1), whether their realizers are really different (in criterion 2), which of the differences are trivial vs. relevant (in criterion 4), etc. The problem of kind-splitting still arises, in which one side says that a purported "kind" is really different subkinds ("split and eliminate:" Aizawa, 2013). For example, P and S (pp. 99–104) used kind-splitting to say that the purported kind, memory, is really many different kinds, such as declarative memory, skills memory, motor learning, and associative learning—to which we retort that all these subtypes of memory involve storage and recall, making them one kind after all-and so on. As another example of the persisting difficulties, if someone claims that two realizers differ (e.g., bird and mammal brains), then it is easy to object by saying they are fundamentally similar. We will apply P and S's valuable Recipe to various cases and handle such difficulties the best we can.

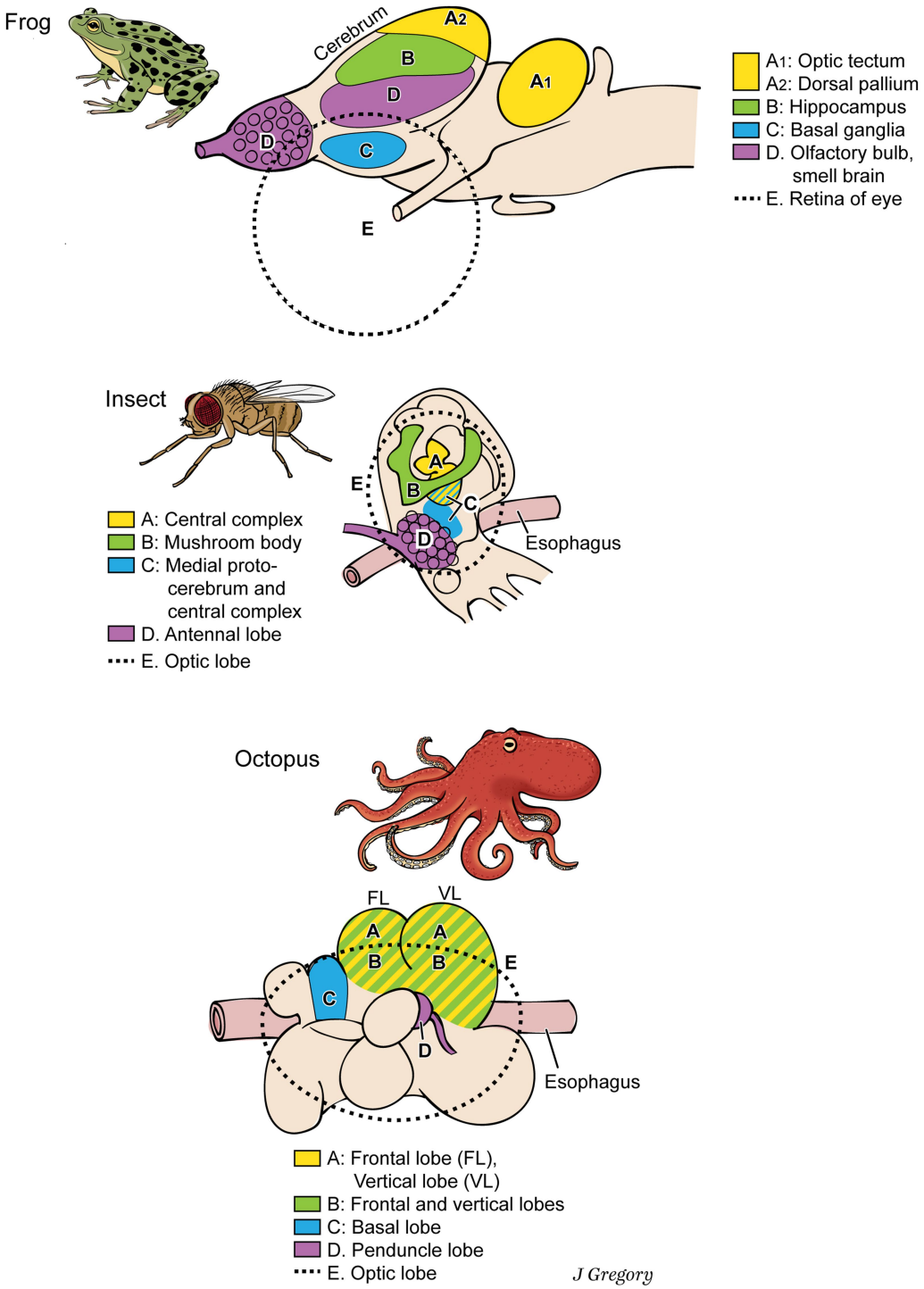


FIGURE 3 | Dissimilar brains of three different taxa of animals with consciousness. The areas with similar functions are colored the same in the different brains. The general code is: **(A)**, image-based consciousness; **(B)**, memory; **(C)**, pre-motor center; **(D)**, smell processing; and **(E)**, visual processing. From *Consciousness Demystified*, MIT, 2018, reproduced with the permission of the copyright holder Mount Sinai Health System.

Neural Plasticity and MR

The argument that is most commonly and traditionally used to support MR is neural plasticity, such as when the functions of a damaged part of the cerebral cortex are taken up over time by other parts of the cortex (Block and Fodor, 1972). P and S questioned two prominent experiments that were said to show multiple realization through neural plasticity (pp. 90–98). First was an experiment by Von Melchner et al. (2000), who

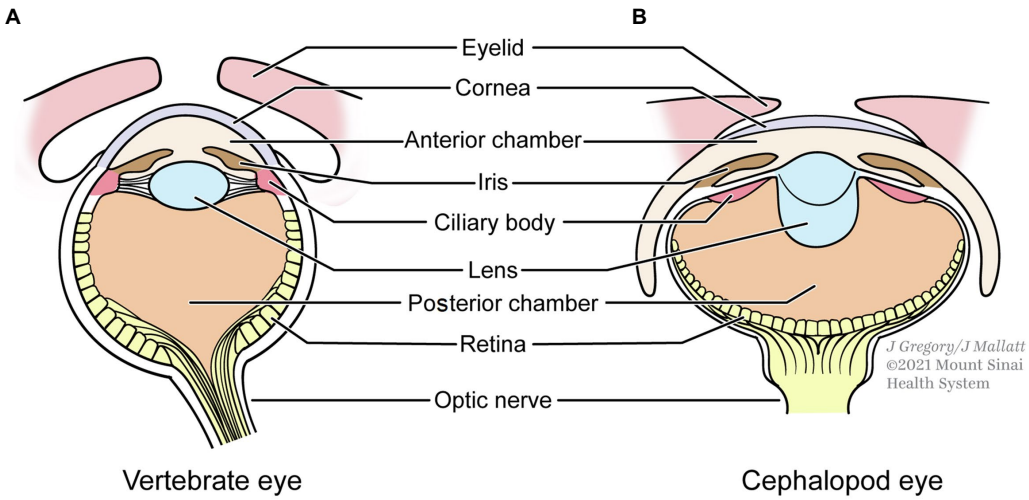


FIGURE 4 | The independently evolved eyes of vertebrates **(A)** and cephalopods **(B)** show strong similarities. **(B)** is redrawn from Hanke and Kelber (2020).

directed the still-developing visual pathway of newborn ferrets away from the usual, visual, cortex to the differently organized *auditory* cortex and found that the "'rewired' ferrets respond as though they perceive stimuli (i.e., light) to be visual rather than auditory." This would be MR because the ferrets had gained a same kind (vision) through a different route that involved the auditory cortex. However, as P and S point out, tests showed the ferrets' vision was degraded, with a diminished discriminatory capacity. Therefore, the normal and rewired visual kinds were not the same, the example fails criterion (1) of the Official Recipe, and this is not MR. We agree with P and S's refutation here. Our disagreements start with the next example.

The second plasticity-related example of multiple realization that P and S sought to refute involves the cerebral cortex of the owl monkey, specifically the part of the somatosensory area that represents the fingers for touch sensation (Merzenich et al., 1983a,b; Kaas, 1991). The experiments showed that cutting the nerve to the ventral, fingerprint, side of the first two fingers, which removed all sensory inputs to the cortical representation of this ventral-finger area, was followed by a plastic reorganization of that brain area so it then processed input from the dorsal, fingernail, side (**Figure 5**). P and S concluded this plasticity does *not* indicate MR, because the ventral-digital area took on a *new* function (of dorsal innervation) and therefore it violated their criterion (1) that says the functional kind must be the same in the two situations, before and after. However, we argue that the experiment does support MR, if we simply shift our perspective over to the dorsal sides of the digits. That is, the sensory processing of this dorsum remains the same functional/mental kind (it is still for touch perception), but now a different cortical-processing area has been added (the area formerly for the ventral sides of the fingers) to the original dorsal processing area. That yields two different realizer areas for the same mental kind, just as MR demands. Therefore, this example of neural plasticity (**Figure 5**) fits MR.

Compensatory Differences and Multiple Realization

Kenneth Aizawa (2013) introduced an argument for multiple realization that he called *multiple realization by compensatory differences* or MRCD. His argument is that when a set of realizing properties contribute jointly to a phenomenon, then changes in some of the properties can be offset by (compensated by) changes in the other properties to keep yielding the original phenomenon. To illustrate this argument, he used equations and formulas for scientific laws as an analogy. Electrical resistance (R) in a wire, for example, is given by $R = l \bullet \rho / A$, where l is the wire's length, ρ is the resistivity of the material that makes up the wire, and A is the wire's cross-sectional area. Thus, the same resistance (kind "R") results if the area (A) is made smaller and this is counterbalanced by a shorter length or else by replacing the wire with one made of a material with a lower ρ. Other examples are Newton's second law of motion, $F = m \bullet a$, where a given force can be attained by a change in mass that balances a change in acceleration, or *vice-versa*, and Ohm's law for an electrical circuit, $I = V/R$, where a given current I can be maintained by a change in voltage V that counterbalances a change in resistance R. Aizawa's MRCD both demonstrates that "there is more than one way to skin a cat" and offers an easily understood reason for this MR thesis.

P and S only briefly addressed the MRCD concept, in a short footnote on page 72 of their book. They argued against MRCD by referring to the $R = l \bullet \rho / A$ example and saying, "In our view, however, these are not multiple realizers of resistance, they are all resisters in the same way." In other words, they are similarly realized, with the reasoning apparently being that the same three compensating variables (l, ρ, and A) vary along gradients, making them one continuum. P and S seem to be saying that MR requires qualitative, not merely quantitative, differences between its realizers.

For us, this argument against MRCD breaks down when the variables have extremely low or high values, and it breaks

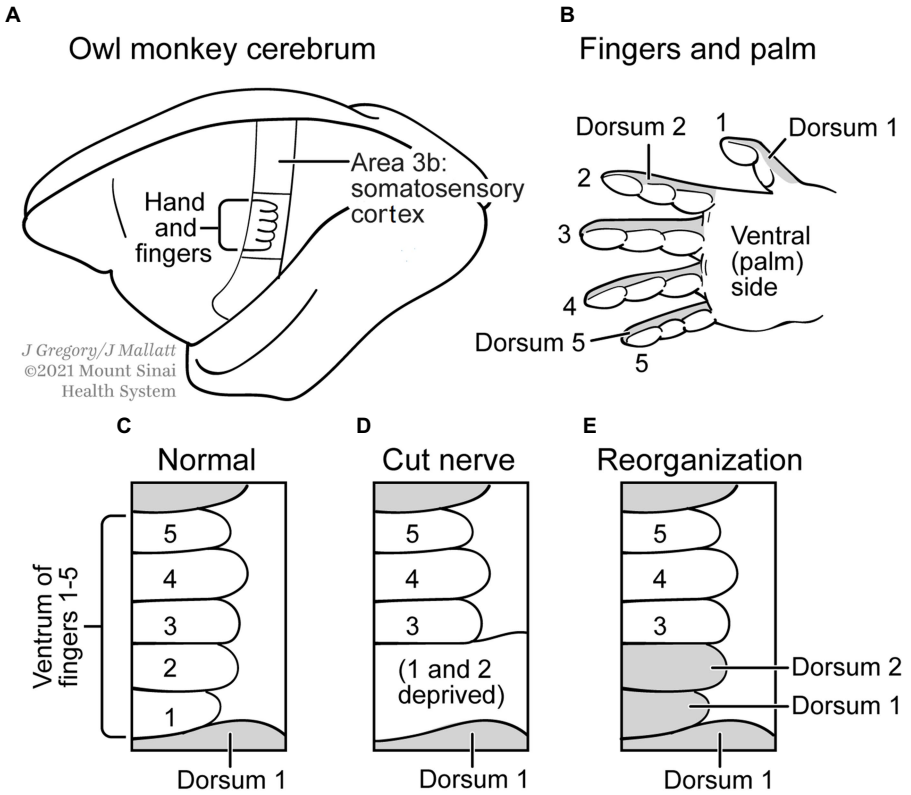


FIGURE 5 | Neural plasticity and multiple realization. **(A)** Cerebral cortex of an owl monkey has an Area S3b, which processes somatosensory (touch) signals from a nerve to the palm side of the hand **(B)**. **(C)** Enlargement of the representation in S3b of a normal monkey, for the five fingers; finger 1 is the thumb and finger 2 is the index finger. **(D)** The representation after the nerve to the first two fingers was cut. **(E)** The areas about a month after the nerve was cut, when some regeneration has occurred. Now the areas for fingers 1 and 2 receive sensory input from the other, *dorsal* (fingernail) side of these digits. Modified from Kaas (1991).

down for practical reasons about physical design. Take the $F = m \bullet a$ example. When the particular force is to be achieved by a huge mass that accelerates and moves slowly, such as an earthmover that crawls along, many of the design concerns are about building a massive motor vehicle; but when that same F is to be achieved by rapidly accelerating a tiny object, such as firing a bullet, then the design concerns are much different, mostly about building a handgun. Thus, the mechanisms behind the realizers are qualitatively different and this is still multiple realization. As another example, take Ohm's law where a particular current I is to be achieved by high voltage V and moderate resistance R. For this, the design can use a powerful lithium battery and an ordinary copper wire. But if the same current I is to be achieved another way—by moderate voltage and low resistance—the design uses an ordinary alkaline battery and a superconducting wire. Again, it is the same realized kind in both cases, they have qualitatively different realizers, and multiple realization (MRCD) is the correct description.

Aizawa's examples involved simple physical states and he had to assume that compensatory differences also characterize the complex brain states with which classical MR questions deal. This assumption is very difficult to test because of the almost universal lack of knowledge of "exactly what the realizers of psychological states are and how they work" (Aizawa, 2013, p. 79). We can, however, offer an apparent example of a multiply-realized compensatory difference that is, though not of a mental state, at least a brain-signaled behavior. This example is the fast way that squids and fish escape through the water when threatened with danger (**Figure 6**). Squids use rapidly conducting giant axons to jet-propel away, whereas fish use rapidly conducting Mauthner axons to bend their body then swim off fast (Shapiro, 2004, p 133; Castelfranco and Hartline, 2016). We consider the escape responses of both animals to be the same "kind," molded by natural selection for survival under the same, threatening, circumstances. Both the types of axons maximize their speeds of impulse conduction but through compensatory differences. For these differences, consider the formula for the propagation velocity (V) of the action potential along the axon that carries the escape signal:

$$V \propto \left(1 / C_m\right) \bullet \left(d / 4 R_m R_i\right)^{1/2}$$

where C_m is the axonal membrane's capacitance, d is the axon's diameter, R_m is the membrane resistance, and R_i is the resistance of the axon's cytoplasm. The squid giant axon increases the V by maximizing the axon's diameter d (to 1–1.5 mm). The fish axon, by contrast, has a coat of myelin that alters both

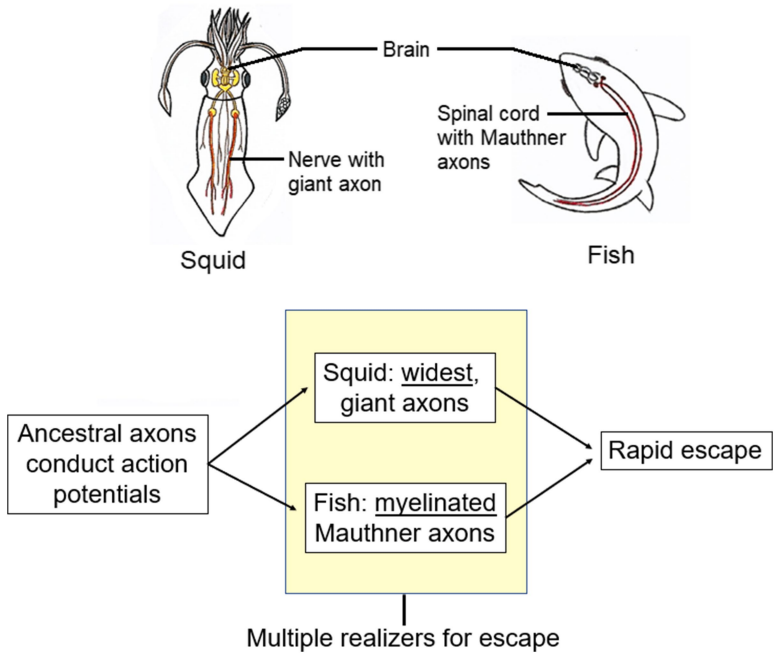


FIGURE 6 | Multiple realizers for signaling rapid escape in squid vs. fish. Squid picture is from Feinberg and Mallatt (2020). Reproduced with the permission of the copyright holder Mount Sinai Health System.

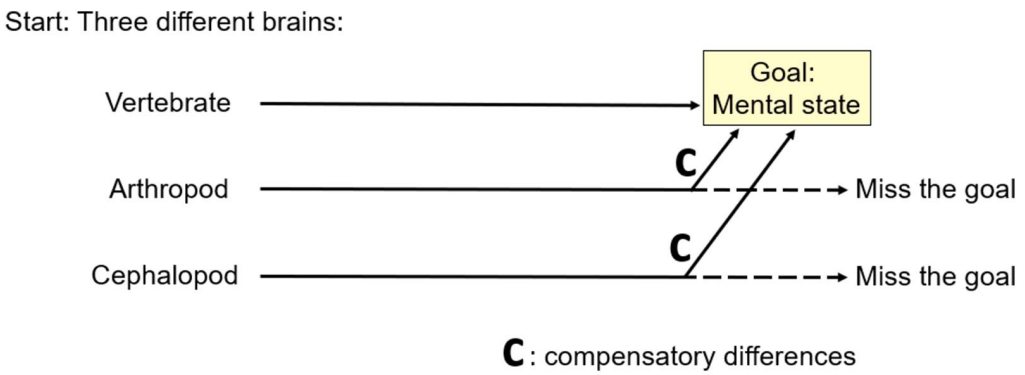


FIGURE 7 | A theoretical reason for multiple realizability by compensatory differences. If different animal lineages start from different places (left), then all but one must evolve compensatory differences if all are to reach a common goal (right).

C_m and R_m in a way that increases V with only a small increase in d (to 0.04–0.09 mm). This is a multiple realization of the function "fast propagation" through a compensatory difference, with squid relying on axonal widening and fish relying more on myelination.

To us, Aizawa's MRCD is convincing because, given evolutionary considerations, it seems like it *must* happen. Here is why (**Figure 7**). As mentioned above, phylogenetic reconstruction indicates the common ancestor of the vertebrates, arthropods, and cephalopods was brainless (Northcutt, 2012) and the immediate ancestors of these three clades had different brains (e.g., Lacalli, 2008; Strausfeld, 2012). Starting from different places demands that MRCD occurred by definition, because otherwise two of the three clades would have missed the goal of the mental state that we argued they do have.

Bird and Mammal Pallia

P and S examined another test case for whether MR exists, comparing the enlarged cerebral pallia of mammals and birds. In mammals, this brain region is dominated by the cerebral cortex (neocortex), and in birds by functionally equivalent regions called the dorsal ventricular ridge (DVR) plus the cortex-like Wulst (**Figure 8**). However, the pallium enlarged independently in birds and mammals, from a smaller and more simply organized pallium in their reptile-like common

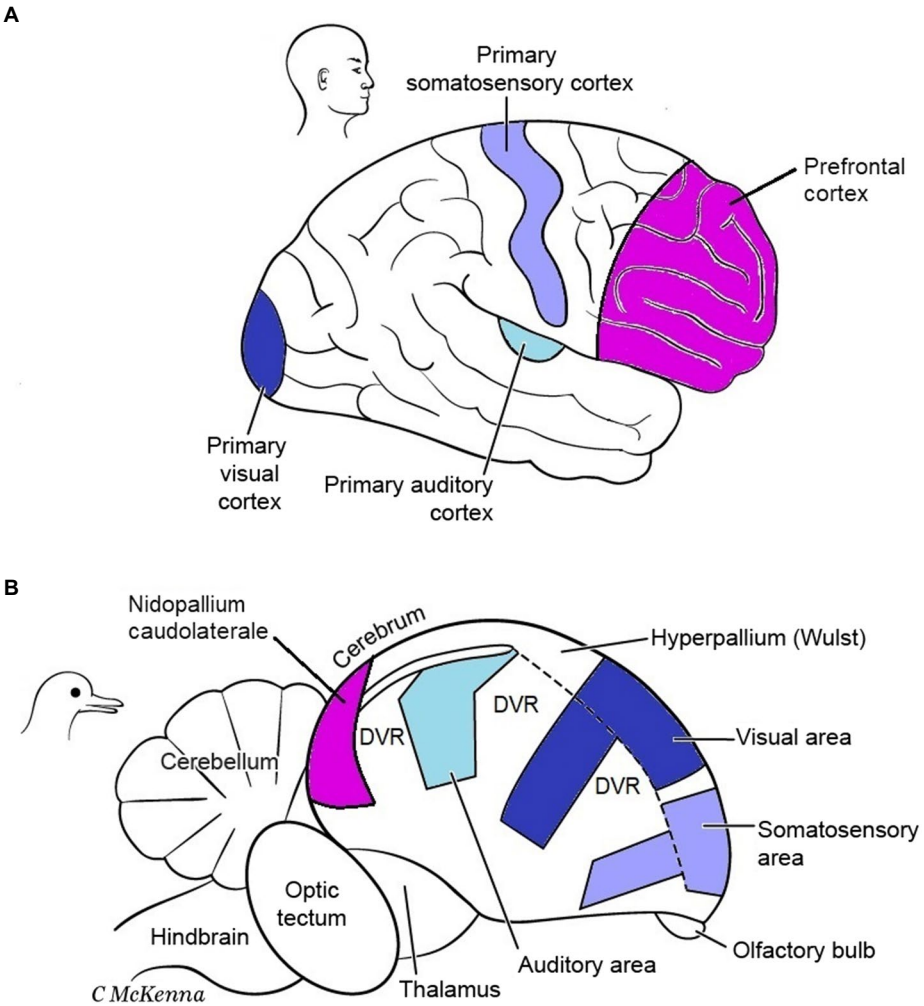


FIGURE 8 | Cerebrums of a mammal **(A)** and bird **(B)**. Functional areas involved in conscious sensory perception and cognition are color-coded. The same functional areas have evolved in different relative locations in these brains. DVR, dorsal ventricular ridge of bird. Modified from Figure 6.9 in Feinberg and Mallatt (2016). The images are reproduced with the permission of the copyright holder Mount Sinai Health System.

ancestor that lived 350 million years ago (Striedter and Northcutt, 2020). Birds and mammals perform many of the same mental tasks, and it is widely accepted that their convergent pallial expansions permitted the higher mental functions that these taxa share, such as more cognitive abilities, increased memory of objects and events, better problem-solving skills, and improved sensory processing for primary consciousness (Feinberg and Mallatt, 2016; Briscoe and Ragsdale, 2018a; Nieder et al., 2020; Tosches, 2021). Like us, P and S consider these functions to be the same "mental kinds" in birds and mammals because on p. 115 they favorably quoted Karten's (2013) characterization of these as "virtually identical outcomes."

Do the neural realizers of these mental kinds differ enough between birds and mammals to indicate MR, or are they similar enough to refute MR instead? As with all MR questions, the detailed neural circuits are not known well enough to answer these questions definitively. However, these are intensely studied brains about which much *is* known, including the basic circuits and many of the differences and similarities (Jarvis et al., 2013; Dugas-Ford and Ragsdale, 2015; Briscoe and Ragsdale, 2018b; Striedter and Northcutt, 2020; Colquitt et al., 2021). Thus, an up-to-date analysis should at least suggest an answer.

Gross structural and functional differences seem to support MR (**Figure 8**). The corresponding functional areas, independently evolved, have different locations in the mammal vs. bird pallia. First, notice the different relative positions of the primary auditory, visual, and somatosensory areas for conscious sensation. Next, notice that the integrative areas for high-level cognition—the prefrontal cortex in mammals and the nidopallium caudolaterale in birds—are in opposite poles of the pallium, front vs. back, respectively (Güntürkün and Bugnyar, 2016). Additionally, mammals have no structure like the DVR of birds. Furthermore, the bird analogues of the six layers of the mammalian cortex are spread widely through the pallium as nuclei (unlayered neuron clusters) or as thick bands (I–VI in **Figure 9**); this bird state is so unlike the

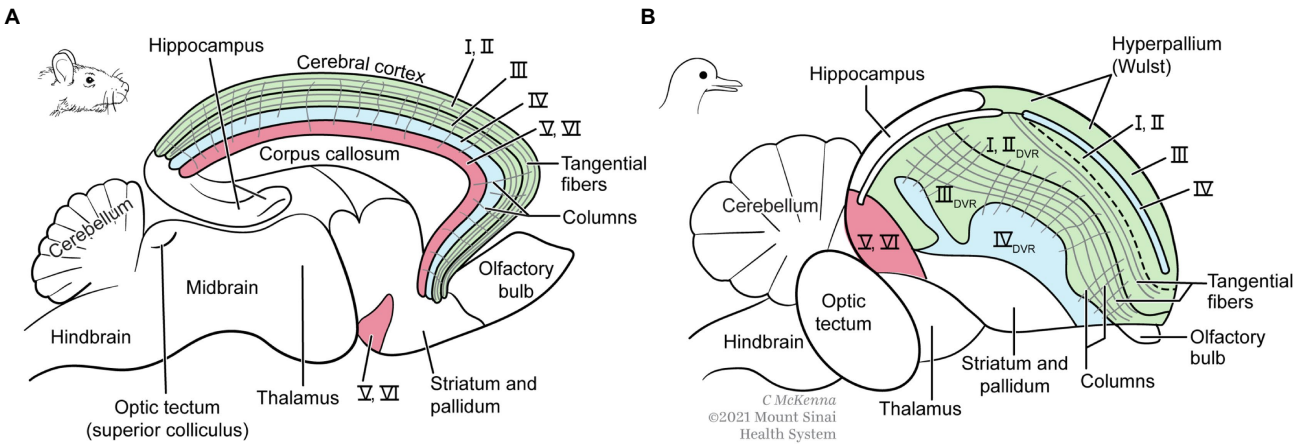


FIGURE 9 | Finer structure of mammal **(A)** and bird **(B)** cerebrums. The mammal brain is cut in half in the sagittal midline. Whereas the mammal neocortex has six thin layers of neurons (I–VI), the analogous structures in birds are thicker and distributed more widely including in the DVR, which mammals lack. Both brains, however, have comparable columns of radially oriented fibers and groups of tangentially oriented fibers. Modified from Figure 6.9 in Feinberg and Mallatt (2016). The images are reproduced with the permission of the copyright holder Mount Sinai Health System.

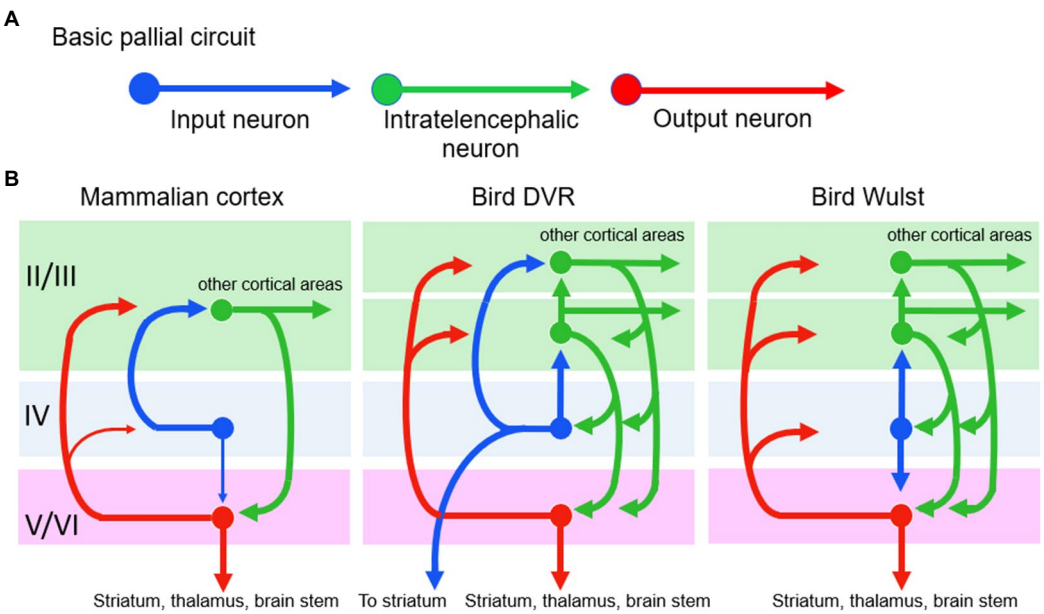


FIGURE 10 | The basic pallial circuit of three kinds of neurons **(A)** is present in mammals and birds. However, differences appear (among the similarities) when the circuit is shown in more detail **(B)**. At left, the numbers II to VI are the numbered layers and structures of **Figure 9**. Modified from Stacho et al. (2020).

mammal state that it took neurobiologists over a century to even identify the comparable regions (Dugas-Ford et al., 2012; Jarvis et al., 2013). Finally, in embryonic mammals, the cortical layering develops in an outside-in sequence unlike that in birds or any other vertebrate (Tosches et al., 2018; Striedter and Northcutt, 2020, p. 390). So far this looks like very different pallial structures causing similar mental states, apparently an overwhelming argument for MR.

Now let us consider P and S's argument *against* this being a case of MR. They declare, after Karten, that the basic pallial circuitry is the same in mammals and birds, so that is a causal identity for the identical outcomes, meaning no MR. **Figure 10A** shows the basic circuit, with an input neuron, an intratelencephalic neuron, and an output neuron. We accept that this three-neuron circuit is homologous in mammals and birds but we say it is too rudimentary to perform the higher mental functions that are considered here. It is basically a three-neuron reflex arc, and reflexes are not higher functions. Even the lamprey, a tiny-brained jawless fish has this basic pallial circuit without any of the higher cognitive functions of mammals and birds (Suryanarayana et al., 2017). No, the bird and mammal

circuits would have to be identical at a *higher* level than this to be evidence for identity and against MR.

Therefore, we must look up to the next level of processing (**Figure 10B**), namely, to the many connections between the three neurons that begin to reflect higher processing. Although this level does show many connectional similarities in birds and mammals, there are notable differences that preclude identity. One difference, shown in the figure, is that in the bird circuits the intratelencephalic neurons (green) send more extensive feedback to the other two neuron types, especially to the input neurons. Another difference is that in the bird DVR the input neurons (blue) project directly to the brain's striatum, a pre-motor region. These differences could be functionally relevant, especially the striatal projection, because birds make more use of pallial-sensory signals to the striatum than mammals do (Striedter and Northcutt, 2020, p. 318). These signals help the birds to make informed decisions about which motor behaviors to execute in any given context. In summary, we are back to finding differences rather than identity and to finding further support for MR.

Although the evidence so far favors differences and MR, it is important to discuss some additional similarities between the bird and mammal pallia (Wang et al., 2010; Feinberg and Mallatt, 2016; Fernández et al., 2020; Stacho et al., 2020). First, the sensory inputs to both these pallia are arranged according to a body map. Second, the bird pallium contains axon fibers that extend radially and mark out columns that resemble the "cortical columns" of mammals; and third, the bird pallium also contains tangentially running fibers that interconnect distant pallial areas and lie in similar places to such fibers in mammals (**Figure 9**). We discount these three similarities, however, because Karten and P and S (pp. 115–117) demanded that they be homologous in order to support an identity theory, but they are demonstrably *not* homologous. That is, the similarities are analogues that evolved separately in birds and mammals, as evidenced by the fact that they are absent in today's reptile relatives of birds—relatives that reveal the pre-bird pallial condition (Striedter and Northcutt, 2020). The reason these similarities evolved independently during brain enlargement in birds and mammals presumably had to do with shared constraints, namely, the need to increase information-processing in more organized and efficient ways, and to save on the cost of axonal wiring (Kaas, 1997, Shapiro, 2004, p. 130). As analogues, they favor the MR interpretation.

We end this section with our formal argument that the "bird-vs.-mammal" example supports MR, contrary to the claim of P and S. According to the Official Recipe, the higher mental kinds in birds and mammals are the same, meeting its criterion (1). The causal realizers show differences (at many levels), meeting criterion (2). The differences between the bird and mammal circuitries could make the mental kinds the same, which would fit criterion (3). And these differences are probably not trivial but relevant to realizing the higher mental processes, which would fit criterion (4).

More Realizability at Lower Levels?

We have focused on the higher levels of the brain, where we found examples of multiple realizability that had relatively few alternate realizers of mental processes. A possible challenge to this *limited* type of realizability is the possibility of *extensive* realizability at the *lower* levels. That is, as one goes lower in the biological hierarchy (from organ to cells to biomolecules) and encounters more and smaller realizers that could contribute to an overall process, the alternate realizers may become more dissimilar and more numerous. Some examples support this possibility. First, if one goes down to the cell level, one finds a large dissimilarity involving animals called ctenophores. These comb jellies (or sea gooseberries) evolved their nerve cells independently of all the other animals with nervous systems, as revealed by ctenophores' unique set of synaptic neurotransmitters (Moroz and Kohn, 2016). Second, the submicroscopic action potentials on which neuronal signaling depends can be generated in various, dissimilar ways; e.g., by influxes of Na^+ in animals vs. influxes of Ca^{2+} in plant cells (Mallatt et al., 2021). Third, down at the intracellular level, many alternate enzymatic pathways can perform the same metabolic role through multiple realizability, a form of redundancy that aids cellular survival (Wagner, 2014, Chapter 6). A fourth example of more MR at lower levels goes down to the genes: A number of studies have found that different genes can account for the same phenotypic adaptation in different organisms (Natarajan et al., 2016; James et al., 2020; Figure 1 in Pyenson and Marraffini, 2020; Colella et al., 2021). While these are all valid examples of MR to add to our growing list, do they really show that MR is more common at lower levels? Do they take us back to standard MRT, with its "very many" possible realizers?

Probably not, because many counterexamples show *identity* at the lower levels. First, some genetic studies of the parallel evolution of phenotypes reveal "identical mutations fixed independently" (Sackton and Clark, 2019). Second, numerous other lower-level features are the same in all animals. These universally conserved features include: the presence of epithelium and connective tissues; the same, eukaryote cell type with the same suite of cellular organelles; the same 64 codons for the genetic code; and the same four nucleotides of DNA (A, C, G, and T; Ruppert et al., 2004). In these lower-level examples, there is far less variability than we found among brain regions at the higher levels (**Figure 3**), throwing doubt on the entire claim for more realizers at the lower levels. Where they are rigidly conserved, the lower-level features seem to reflect strong stabilizing selection for survival (e.g., epithelial sheets are the most effective tissues for borders in animal bodies; animal cells without all the typical organelles would be less fit). Therefore, whether or not the instances of multiple realization are more numerous at lower levels of the biological hierarchy, they remain limited by survival constraints. Such constraints operate at every level of biological hierarchies and the multiple-constraint part of Shapiro's thesis still holds true.

Conclusion of Part 2

We agree substantially with the ideas of P and S, but not completely. The disagreements are that we accepted more examples of MR than they did (e.g., neural plasticity,

TABLE 3 | Comparison of the theories presented in this paper, on the realizability of mental states in different taxa.

Standard Multiple Realizability (Bickle, 2020)	**Mental Constraint of Shapiro (2004): MCT**	**Modest Identity of Polger and Shapiro (2016): MIT**	**Our Constrained Multiple Realizability**
1. Many realizers for each mental kind (thousands or more)	1. Few realizers for each mental kind (handful)	1–3. Same as for MCT, and rejects most of the classic examples of multiple realization	1. Few realizers for each mental kind (but can be more than a handful)
2. Constraints are not recognized	2. Constraints are common		2. Constraints are common
3. Convergent evolution is not recognized	3. Convergent evolution is important		3. Convergent evolution is important
4. No identity of mind and brain.	4. Mind-brain identity is not refuted by any multiple realizability	4. Promotes a kind of mind-brain identity by saying strong similarities in brain mechanisms are effectively identities; such identities are common, but MIT tolerates at least some instances of multiply-realized non-identities	4. Strong similarities are not identities, so we recognize more examples of true multiple realization than MIT does. Ours is more of a highly constrained version of MRT than an identity theory

bird-mammal pallia, and alternate enzymatic pathways for cell metabolism) and we accept Aizawa's (2013) proposal that compensatory differences generate multiple realizability. Thus, we say that P and S went too far in arguing against MR. We found that there can be more ways to achieve a mental state than just Shapiro's "handful" (though still fewer ways than standard MRT claims). It should be easy to reconcile our disagreements with P and S because they explicitly designed their modest identity theory to allow more instances of true MR, as long as this also allows some substantial instances of identity (p. 34). A central point of agreement is that both we and they recognize the importance of convergent constraints in limiting the number of realizations, which the standard MRT—with its almost numberless realizations—does not.

CONCLUSION

Our consideration of animal evolution reveals that the emergence of consciousness proceeded under many constraints and therefore involved strong evolutionary convergences between vertebrates, arthropods, and cephalopods (**Table 1**), as well as between birds and mammals (**Figures 8**–**10**). This emergence proceeded along the multiple routes of a highly constrained multiple realizability. **Table 3** provides a summary by comparing our present conclusions with the standard MRT, Shapiro's (2004) mental constraint thesis, and Polger and Shapiro's (2016) modest identity theory.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article. Further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

JM analyzed the theories and wrote most of the manuscript. TF provided much of the information on consciousness and emergence. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

Thanks are extended to Jill K. Gregory and Courtney McKenna for help with the figures.

REFERENCES

Aizawa, K. (2013). Multiple realization by compensatory differences. *Eur. J. Philos. Sci.* 3, 69–86. doi: 10.1007/s13194-012-0058-6

Baysan, U. (2019). "Emergence, function and realization," in *Routledge Handbook of Emergence*. eds. S. Gibb, R. F. Hendry and T. Lancaster (New York: Taylor & Francis), 77–86.

Bedau, M. A. (2008). "Downward causation and the autonomy of weak emergence," in *Emergence: Contemporary Readings in Philosophy and Science*. eds. M. A. Bedau and P. Humphreys (Cambridge: MIT Press), 155–188.

Ben-Haim, M. S., Dal Monte, O., Fagan, N. A., Dunham, Y., Hassin, R. R., Chang, S. W., et al. (2021). Disentangling perceptual awareness from nonconscious processing in rhesus monkeys (*Macaca mulatta*). *Proc. Natl. Acad. Sci.* 118:e2017543118. doi: 10.1073/pnas.2017543118

Berridge, K. C., and Kringelbach, M. L. (2015). Pleasure systems in the brain. *Neuron* 86, 646–664. doi: 10.1016/j.neuron.2015.02.018

Betley, J. N., Xu, S., Cao, Z. F. H., Gong, R., Magnus, C. J., Yu, Y., et al. (2015). Neurons for hunger and thirst transmit a negative-valence teaching signal. *Nature* 521, 180–185. doi: 10.1038/nature14416

Bickle, J. (2020). Multiple realizability. Encyclopedia of Cognitive Science. Available at: https://plato.stanford.edu/archives/spr2019/entries/multiple-realizability (Accessed June 24, 2021).

Bishop, R., and Silberstein, M. (2019). "Complexity and feedback," in *Routledge Handbook of Emergence*. eds. S. Gibb, R. F. Hendry and T. Lancaster (New York: Taylor & Francis), 145–156.

Block, N. J., and Fodor, J. A. (1972). What psychological states are not. *Philos. Rev.* 81, 159–181. doi: 10.2307/2183991

Blundell, S. J. (2019). "Phase transitions, broken symmetry, and the renormalization group," in *Routledge Handbook of Emergence*. eds. S. Gibb, R. F. Hendry and T. Lancaster (New York: Taylor & Francis), 237–247.

Briscoe, S. D., and Ragsdale, C. W. (2018a). Homology, neocortex, and the evolution of developmental mechanisms. *Science* 362, 190–193. doi: 10.1126/science.aau3711

Briscoe, S. D., and Ragsdale, C. W. (2018b). Molecular anatomy of the alligator dorsal telencephalon. *J. Comp. Neurol.* 526, 1613–1646. doi: 10.1002/cne.24427

Brodal, P. (2016). *The Central Nervous System: Structure and Function. 5th Edn*. New York: Oxford University Press.

Cabanac, M. (1996). On the origin of consciousness, a postulate and its corollary. *Neurosci. Biobehav. Rev.* 20, 33–40. doi: 10.1016/0149-7634(95)00032-A

Castelfranco, A. M., and Hartline, D. K. (2016). Evolution of rapid nerve conduction. *Brain Res.* 1641, 11–33. doi: 10.1016/j.brainres.2016.02.015

Colella, J. P., Tigano, A., Dudchenko, O., Omer, A. D., Khan, R., Bochkov, I. D., et al. (2021). Limited evidence for parallel evolution among desert-adapted *Peromyscus* deer mice. *J. Hered.* 112, 286–302. doi: 10.1093/jhered/esab009

Colquitt, B. M., Merullo, D. P., Konopka, G., Roberts, T. F., and Brainard, M. S. (2021). Cellular transcriptomics reveals evolutionary identities of songbird vocal circuits. *Science* 371:eabd9704. doi: 10.1126/science.abd9704

Conway Morris, S. (2003). *Life's Solution: Inevitable Humans in a Lonely Universe*. Cambridge: Cambridge University Press.

Dugas-Ford, J., and Ragsdale, C. W. (2015). Levels of homology and the problem of neocortex. *Annu. Rev. Neurosci.* 38, 351–368. doi: 10.1146/annurev-neuro-071714-033911

Dugas-Ford, J., Rowell, J. J., and Ragsdale, C. W. (2012). Cell-type homologies and the origins of the neocortex. *Proc. Natl. Acad. Sci.* 109, 16974–16979. doi: 10.1073/pnas.1204773109

Ellis, G. F. (2012). Top-down causation and emergence: some comments on mechanisms. *Interface Focus* 2, 126–140. doi: 10.1098/rsfs.2011.0062

Eschbach, C., Fushiki, A., Winding, M., Schneider-Mizell, C. M., Shao, M., Arruda, R., et al. (2020a). Recurrent architecture for adaptive regulation of learning in the insect brain. *Nat. Neurosci.* 23, 544–555. doi: 10.1038/s41593-020-0607-9

Eschbach, C., Fushiki, A., Winding, M., Afonso, B., Andrade, I. V., Cocanougher, B. T., et al. (2020b). Circuits for integrating learnt and innate valences in the fly brain. BioRxiv. doi: 10.1101/2020.04.23.058339 (in press).

Feinberg, T. E., and Mallatt, J. (2013). The evolutionary and genetic origins of consciousness in the Cambrian period over 500 million years ago. *Front. Psychol.* 4:667. doi: 10.3389/fpsyg.2013.00667

Feinberg, T. E., and Mallatt, J. (2016). *The Ancient Origins of Consciousness: How the Brain Created Experience*. Cambridge, MA: MIT Press.

Feinberg, T. E., and Mallatt, J. (2018). *Consciousness Demystified*. Cambridge, MA: MIT Press.

Feinberg, T. E., and Mallatt, J. (2019). Subjectivity "demystified": neurobiology, evolution, and the explanatory gap. *Front. Psychol.* 10:1686. doi: 10.3389/fpsyg.2019.01686

Feinberg, T. E., and Mallatt, J. (2020). Phenomenal consciousness and emergence: eliminating the explanatory gap. *Front. Psychol.* 11:1041. doi: 10.3389/fpsyg.2020.01041

Felsenberg, J., Barnstedt, O., Cognigni, P., Lin, S., and Waddell, S. (2017). Re-evaluation of learned information in *Drosophila*. *Nature* 544, 240–244. doi: 10.1038/nature21716

Fernández, M., Ahumada-Galleguillos, P., Sentis, E., Marín, G., and Mpodozis, J. (2020). Intratelencephalic projections of the avian visual dorsal ventricular ridge: Laminarly segregated, reciprocally and topographically organized. *J. Comp. Neurol.* 528, 321–359. doi: 10.1002/cne.24757

Güntürkün, O., and Bugnyar, T. (2016). Cognition without cortex. *Trends Cogn. Sci.* 20, 291–303. doi: 10.1016/j.tics.2016.02.001

Hanke, F. D., and Kelber, A. (2020). The eye of the common octopus (*Octopus vulgaris*). *Front. Physiol.* 10:1637. doi: 10.3389/fphys.2019.01637

Hartline, H. K., Wagner, H. G., and Ratliff, F. (1956). Inhibition in the eye of Limulus. *J. Gen. Physiol.* 39, 651–673. doi: 10.1085/jgp.39.5.651

James, M. E., Wilkinson, M. J., North, H. L., Engelstädter, J., and Ortiz-Barrientos, D. (2020). A framework to quantify phenotypic and genotypic parallel evolution. BioRxiv. doi: 10.1101/2020.02.05.936450 (in press).

Jarvis, E. D., Yu, J., Rivas, M. V., Horita, H., Feenders, G., Whitney, O., et al. (2013). Global view of the functional molecular organization of the avian cerebrum: mirror images and functional columns. *J. Comp. Neurol.* 521, 3614–3665. doi: 10.1002/cne.23404

Kaas, J. H. (1991). Plasticity of sensory and motor maps in adult mammals. *Annu. Rev. Neurosci.* 14, 137–167. doi: 10.1146/annurev.ne.14.030191.001033

Kaas, J. H. (1997). Topographic maps are fundamental to sensory processing. *Brain Res. Bull.* 44, 107–112. doi: 10.1016/S0361-9230(97)00094-4

Kaiser, M., Goerner, M., and Hilgetag, C. C. (2007). Criticality of spreading dynamics in hierarchical cluster networks without inhibition. *New J. Phys.* 9:110. doi: 10.1088/1367-2630/9/5/110

Kandel, E., Koester, J.D., Mack, S.H., and Siegelbaum, S. (2021). *Principles of Neuroscience. 6th Edn*. New York: McGraw-Hill.

Karten, H. J. (2013). Neocortical evolution: neuronal circuits arise independently of lamination. *Curr. Biol.* 23, R12–R15. doi: 10.1016/j.cub.2012.11.013

Kim, J. (2008). "The nonreductivist's troubles with mental causation," in *Emergence: Contemporary Readings in Philosophy and Science*. eds. M. A. Bedau and P. Humphreys (Cambridge, MA: MIT Press), 427–445.

Lacalli, T. C. (2008). Basic features of the ancestral chordate brain: a protochordate perspective. *Brain Res. Bull.* 75, 319–323. doi: 10.1016/j.brainresbull.2007.10.038

Macdonald, C., and Macdonald, G. (2019). "Emergence and non-reductive physicalism," in *Routledge Handbook of Emergence*. eds. S. Gibb, R. F. Hendry and T. Lancaster (New York: Taylor & Francis), 195–205.

Mallatt, J. (2021a). A traditional scientific perspective on the integrated information theory of consciousness. *Entropy* 23:650. doi: 10.3390/e23060650

Mallatt, J. (2021b). Unlimited associative learning and consciousness: further support and some caveats about a link to stress. *Biol. Philos.* 36:22. doi: 10.1007/s10539-021-09798-y

Mallatt, J., Blatt, M. R., Draguhn, A., Robinson, D. G., and Taiz, L. (2021). Debunking a myth: plant consciousness. *Protoplasma* 258, 459–476. doi: 10.1007/s00709-020-01579-w

Mallatt, J., and Feinberg, T. E. (2020). Sentience in evolutionary context. *Anim. Sentience* 5, 1–5. doi: 10.51291/2377-7478.1599

McGhee, G. R. (2019). *Convergent Evolution on Earth: Lessons for the Search for Extraterrestrial Life*. Cambridge, MA: MIT Press.

McLeish, T. (2019). "Soft matter—an emergent interdisciplinary science of emergent entities," in *Routledge Handbook of Emergence*. eds. S. Gibb, R. F. Hendry and T. Lancaster (New York: Taylor & Francis), 248–264.

Mengistu, H., Huizinga, J., Mouret, J. B., and Clune, J. (2016). The evolutionary origins of hierarchy. *PLoS Comput. Biol.* 12:e1004829. doi: 10.1371/journal.pcbi.1004829

Merzenich, M. M., Kaas, J. H., Wall, J., Nelson, R. J., Sur, M., and Felleman, D. (1983a). Topographic reorganization of somatosensory cortical areas 3b and 1 in adult monkeys following restricted deafferentation. *Neuroscience* 8, 33–55. doi: 10.1016/0306-4522(83)90024-6

Merzenich, M. M., Kaas, J. H., Wall, J. T., Sur, M., Nelson, R. J., and Felleman, D. J. (1983b). Progression of change following median nerve section in the cortical representation of the hand in areas 3b and 1 in adult owl and squirrel monkeys. *Neuroscience* 10, 639–665.

Moroz, L. L., and Kohn, A. B. (2016). Independent origins of neurons and synapses: insights from ctenophores. *Philos. Trans. Royal Soc. B Biol. Sci.* 371:20150041. doi: 10.1098/rstb.2015.0041

Nagel, T. (1974). What is it like to be a bat? *Philos. Rev.* 83, 435–450. doi: 10.2307/2183914

Nahmad-Rohen, L., and Vorobyev, M. (2019). Contrast sensitivity and behavioural evidence for lateral inhibition in octopus. *Biol. Lett.* 15:20190134. doi: 10.1098/rsbl.2019.0134

Namburi, P., Al-Hasani, R., Calhoon, G. G., Bruchas, M. R., and Tye, K. M. (2016). Architectural representation of valence in the limbic system. *Neuropsychopharmacology* 41, 1697–1715. doi: 10.1038/npp.2015.358

Natarajan, C., Hoffmann, F. G., Weber, R. E., Fago, A., Witt, C. C., and Storz, J. F. (2016). Predictable convergence in hemoglobin function has unpredictable molecular underpinnings. *Science* 354, 336–339. doi: 10.1126/science.aaf9070

Nieder, A., Wagener, L., and Rinnert, P. (2020). A neural correlate of sensory consciousness in a corvid bird. *Science* 369, 1626–1629. doi: 10.1126/science.abb1447

Northcutt, R. G. (2012). Evolution of centralized nervous systems: two schools of evolutionary thought. *Proc. Natl. Acad. Sci.* 109(Suppl. 1), 10626–10633. doi:10.1073/pnas.1201889109

Panksepp, J. (2016). The cross-mammalian neurophenomenology of primal emotional affects: From animal feelings to human therapeutics. *J. Comp. Neurol.* 524, 1624–1635. doi: 10.1002/cne.23969

Perry, C. J., and Chittka, L. (2019). How foresight might support the behavioral flexibility of arthropods. *Curr. Opin. Neurobiol.* 54, 171–177. doi: 10.1016/j.conb.2018.10.014

Place, U. T. (1956). Is consciousness a brain process? *Br. J. Psychol.* 47, 44–50. doi: 10.1111/j.2044-8295.1956.tb00560.x

Polger, T. W., and Shapiro, L. A. (2016). *The Multiple Realization Book*. Oxford: Oxford University Press.

Putnam, H. (1967). Psychological predicates. *Art Mind Religion* 1, 37–48.

Pyenson, N. C., and Marraffini, L. A. (2020). Co-evolution within structured bacterial communities results in multiple expansion of CRISPR loci and enhanced immunity. *Elife* 9:e53078. doi: 10.7554/eLife.53078

Rubinov, M., Sporns, O., Thivierge, J. P., and Breakspear, M. (2011). Neurobiologically realistic determinants of self-organized criticality in networks of spiking neurons. *PLoS Comput. Biol.* 7:e1002038. doi: 10.1371/journal.pcbi.1002038

Ruppert, E. E., Barnes, R. D., and Fox, R. S. (2004). *Invertebrate Zoology: A Functional Evolutionary Approach*. Belmont, CA: Thomson: Brooks/Cole.

Sackton, T. B., and Clark, N. (2019). Convergent evolution in the genomics era: new insights and directions. *Phil. Trans. R. Soc. B* 374:20190102. doi: 10.1098/rstb.2019.0102

Sanes, J. R., and Zipursky, S. L. (2010). Design principles of insect and vertebrate visual systems. *Neuron* 66, 15–36. doi: 10.1016/j.neuron.2010.01.018

Seth, A. (2009). "Functions of consciousness," in *Elsevier Encyclopedia of Consciousness*. ed. W. P. Banks (San Francisco: Elsevier), 279–293.

Shapiro, L.A. (2004). *The Mind Incarnate*. Cambridge, MA: MIT Press.

Shapiro, L. A. (2008). How to test for multiple realization. *Philos. Sci.* 75, 514–525. doi: 10.1086/594503

Shapiro, L. A., and Polger, T. W. (2012). "Identity, variability, and multiple realization in the special sciences," in *New Perspectives on Type Identity: The Mental and the Physical*. eds. S. Gozzano and C. Hill (Cambridge: Cambridge University Press), 264–288.

Shigeno, S., Andrews, P. L., Ponte, G., and Fiorito, G. (2018). Cephalopod brains: an overview of current knowledge to facilitate comparison with vertebrates. *Front. Physiol.* 9:952. doi: 10.3389/fphys.2018.00952

Siju, K. P., Štih, V., Aimon, S., Gjorgjieva, J., Portugues, R., and Kadow, I. C. G. (2020). Valence and state-dependent population coding in dopaminergic neurons in the fly mushroom body. *Curr. Biol.* 30, 2104–2115. doi: 10.1016/j.cub.2020.04.037

Simon, H. A. (2002). Near decomposability and the speed of evolution. *Ind. Corp. Chang.* 11, 587–599. doi: 10.1093/icc/11.3.587

Smart, J. J. (1959). Sensations and brain processes. *Philos. Rev.* 68, 141–156. doi: 10.2307/2182164

Solms, M. (2019). The hard problem of consciousness and the free energy principle. *Front. Psychol.* 9:2714. doi: 10.3389/fpsyg.2018.02714

Sporns, O., and Betzel, R. F. (2016). Modular brain networks. *Annu. Rev. Psychol.* 67, 613–640. doi: 10.1146/annurev-psych-122414-033634

Stacho, M., Herold, C., Rook, N., Wagner, H., Axer, M., Amunts, K., et al. (2020). A cortex-like canonical circuit in the avian forebrain. *Science* 369:eabc5534. doi: 10.1126/science.abc5534

Strausfeld, N. J. (2012). *Arthropod Brains: Evolution, Functional Elegance, and Historical Significance*. Cambridge, MA: Belknap Press of Harvard University Press.

Striedter, G.F., and Northcutt, R.G. (2020). *Brains Through Time*. Oxford: Oxford University Press.

Suryanarayana, S. M., Robertson, B., Wallén, P., and Grillner, S. (2017). The lamprey pallium provides a blueprint of the mammalian layered cortex. *Curr. Biol.* 27, 3264–3277. doi: 10.1016/j.cub.2017.09.034

Tahko, T. E. (2020). Where do you get your protein? Or: biochemical realization. *Br. J. Philos. Sci.* 71, 799–825. doi: 10.1093/bjps/axy044

Tosches, M. A. (2021). Different origins for similar brain circuits. *Science* 371, 676–677. doi: 10.1126/science.abf9551

Tosches, M. A., Yamawaki, T. M., Naumann, R. K., Jacobi, A. A., Tushev, G., and Laurent, G. (2018). Evolution of pallium, hippocampus, and cortical cell types revealed by single-cell transcriptomics in reptiles. *Science* 360, 881–888. doi: 10.1126/science.aar4237

Tye, K. M. (2018). Neural circuit motifs in valence processing. *Neuron* 100, 436–452. doi: 10.1016/j.neuron.2018.10.001

Vogel, S. (1998). *Cats' Paws and Catapults: Mechanical Worlds of Nature and People*. New York: WW Norton & Company.

Von Melchner, L., Pallas, S. L., and Sur, M. (2000). Visual behaviour mediated by retinal projections directed to the auditory pathway. *Nature* 404, 871–876. doi: 10.1038/35009102

Wagner, A. (2014). *Arrival of the Fittest: Solving Evolution's Greatest Puzzle*. New York: Penguin.

Wang, Y., Brzozowska-Prechtl, A., and Karten, H. J. (2010). Laminar and columnar auditory cortex in avian brain. *Proc. Natl. Acad. Sci.* 107, 12676–12681. doi: 10.1073/pnas.1006645107

Wang, Z. Y., and Ragsdale, C. W. (2019). "Cephalopod nervous system organization," in *Oxford Research Encyclopedia of Neuroscience*. ed. P. Katz (Oxford: Oxford University Press).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Systems Neuroscience

HYPOTHESIS AND THEORY
published: 20 October 2021
doi: 10.3389/fnsys.2021.697129

Consciousness as a Product of Evolution: Contents, Selector Circuits, and Trajectories in Experience Space

*Thurston Lacalli**

Biology Department, University of Victoria, Victoria, BC, Canada

Conscious experience can be treated as a complex unified whole, but to do so is problematic from an evolutionary perspective if, like other products of evolution, consciousness had simple beginnings, and achieved complexity only secondarily over an extended period of time as new categories of subjective experience were added and refined. The premise here is twofold, first that these simple beginnings can be investigated regardless of whether the ultimate source of subjective experience is known or understood, and second, that of the contents known to us, the most accessible for investigation will be those that are, or appear, most fundamental, in the sense that they resist further deconstruction or analysis. This would include qualia as they are usually defined, but excludes more complex experiences (here, formats) that are structured, or depend on algorithmic processes and/or memory. Vision and language for example, would by this definition be formats. More formally, qualia, but not formats, can be represented as points, lines, or curves on a topological experience space, and as domains in a configuration space representing a subset of neural correlates of consciousness, the selector circuits (SCs), responsible for ensuring that a particular experience is evoked rather than some other. It is a matter of conjecture how points in SC-space map to experience space, but both will exhibit divergence, insuring that a minimal distance separates points in experience space representing different qualia and the SCs that evoke them. An analysis of how SCs evolve over time is used to highlight the importance of understanding patterns of descent among putative qualia, i.e., their homology across species, and whether this implies descent from an ancestral experience, or ur-quale, that combines modes of experience that later came to be experienced separately. The analysis also provides insight into the function of consciousness as viewed from an evolutionary perspective, defined here in terms of the access it allows to regions of SC-space that would otherwise be unavailable to real brains, to produce consciously controlled behaviors that could otherwise not occur.

Keywords: qualia versus formats as contents, neural correlates of consciousness, neural algorithms, topological representations, configuration spaces

OPEN ACCESS

Edited by:
Raphael Fernandes Casseb, State University of Campinas, Brazil

Reviewed by:
Joseph Neisser, Grinnell College, United States
Jon Mallatt, Washington State University, United States

***Correspondence:**
Thurston Lacalli
lacalli@uvic.ca

Received: *18 April 2021*
Accepted: *24 September 2021*
Published: *20 October 2021*

Citation:
Lacalli T (2021) Consciousness as a Product of Evolution: Contents, Selector Circuits, and Trajectories in Experience Space. Front. Syst. Neurosci. 15:697129. doi: 10.3389/fnsys.2021.697129

INTRODUCTION

Investigating the nature of consciousness is tricky exercise, a good part of which revolves around the hard problems and explanatory gaps beloved of philosophers (Levine, 1983, 2009; Chalmers, 1995; Van Gulick, 2018). This account is less concerned with those issues, i.e., consciousness as a phenomenon, than with the nature of consciousness as a product of evolution. More specifically, the issue here is a practical one, of finding a conceptual framework for dealing with the action of natural selection on the neural circuits that underpin conscious experience (here, by convention, simply "experience"), and how changes to the circuitry change the experience. How neural circuits evolve is a complex issue in its own right (Tosches, 2017). Adding consciousness to the mix is even more problematic, and perhaps uniquely so, in that we have no way as yet to identify the neural circuits responsible for evoking conscious sensations, and no way beyond inference to assess consciousness in taxa other than our own. But there is no justification for supposing *a priori* that a systematic reductionist approach will not eventually succeed in unraveling the mysteries of consciousness as it has with so many other natural phenomena.

Complex systems of interacting components clearly can have unexpected properties with the potential to provide a source for evolutionary innovation (Solé and Valverde, 2020), and this feature has been used to advantage in a number of theories of consciousness, including integrated information theory and some variants of computational, global workspace and higher order theories (Dehaene and Naccache, 2001; Piccinini and Bahar, 2013; Oizumi et al., 2014; Brown et al., 2019). And indeed, if vertebrate consciousness is entirely a product of cortico-thalamic circuitry, a widely accepted view (Butler, 2008), then complexity would seem to be inextricably linked with consciousness of any kind. Here, in contrast, my assumption is that, like everything else in evolution, complex forms of consciousness are more likely than not to have evolved from simple antecedents that were progressively elaborated and refined over an extended period of evolutionary time in ways that can be understood step by step in adaptive terms. This supposition is receiving increasing attention (Baron and Klein, 2016; Feinberg and Mallatt, 2016; Godfrey-Smith, 2016; Lacalli, 2018), and there is a recognition that even quite early vertebrates may play a role in the story if brain structures evolutionarily older than neocortex are involved, as has been argued for olfactory centers (Shepherd, 2007; Merrick et al., 2014; deVries and Ward, 2016), the optic tectum (Feinberg and Mallatt, 2016), subcortical telencephalic centers, and nuclei in the thalamus and midbrain (Merker, 2005, 2007, 2013; Ward, 2011; Woodruff, 2017). We would then have a much-expanded evolutionary window, materially increasing the prospects of finding vestiges of early stages in the transition from consciousness as it first emerged in evolution to something more complex. The cortex in such scenarios then appears in a different light, as less a precondition for having consciousness of any kind, than a device for exploiting more fully a capability the brain may already have possessed.

What approach should one then take when investigating consciousness from an evolutionary perspective? Consider the skeleton, another complex product of evolution: it consists of diverse parts, each precisely shaped to a purpose and assembled in a way that allows that assemblage to function effectively as a whole. By analogy, the diverse parts from which evolved consciousness is constructed are its distinguishable contents, and the evolutionary questions one can ask about these concern the role each part plays in the whole, and the means by which the whole is coordinated. This presupposes also that the contents of consciousness can be dealt with individually, as entities, and investigated as such. For my purposes I assume this to be the case. Accepting the counterargument, that consciousness is indivisible (e.g., Dainton, 2000; Tye, 2003), leads to a very different analytical focus. From an evolutionary perspective, the unity of consciousness is far more likely to be adaptive rather than intrinsic, in other words a secondary feature, refined progressively and of necessity because no product of evolution is of any use unless its constituent parts operate together in a coordinated way.

The analysis developed here focuses on selected individual contents, and is directed at the question of evolutionary change in general terms, rather than the pros and cons of any particular evolutionary scenario. Issues concerning the hard problems as usually defined are deferred because, from an evolutionary perspective, it is not important what consciousness "is" or from what it originates, only that it is useful (Lacalli, 2020, see Kostic, 2017 for a philosophical justification). As to why consciousness is useful, there will be both specific answers that highlight the relative advantages of conscious decision-making over reflex action in a given behavioral context (Velmans, 2012; Black, 2021), and a general answer that relates to the access consciousness provides, through the evolutionary process, to circuitry variants and behavioral outcomes that could otherwise not exist, as discussed in the concluding section (section "Conclusions, and the Function of Consciousness").

A second set of questions concerns what can be said about the way the neural correlates of consciousness (NCCs) and the sensations they evoke will themselves evolve. These are explored below in a set of thought experiments, using two hypothetical spaces, one for neural circuitry (SC-space, described in the section "Selector Circuits: Robustness and Routes to Innovation") and the other for subjective experience (E-space, described in the section "Trajectories in Experience Space"). The exercise is topological in a general way, with SC-space conceived of as a configuration space (**Figures 1**–**3**) and E-space as its non-physical counterpart (**Figure 4**). This choice limits the analysis to the simplest of contents (as explained in the section "Categorizing Contents") in order to avoid the methodological problems of dealing with sequential processes, which for a topological approach might employ graph theory or recurrent neural networks, the latter being currently a favored model of choice (Schmidhuber, 2015; Yu et al., 2019). The exercise as a whole has practical value given the prospect that, through a combination of innovative optogenetic, 3D reconstruction and electrical recording tools (e.g., Marques et al., 2019; Abbott, 2020), an increasing amount of data relating to NCC activity can be expected in the not-to-distant future. In consequence, it is timely to begin thinking about what such data may reveal,

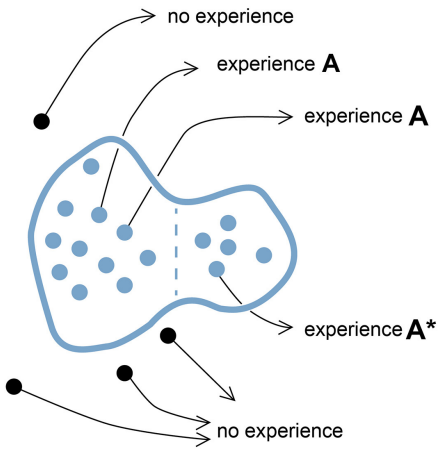


FIGURE 1 | A point cloud in SC-space. The diagram represents a configuration space where distance measures incremental differences in the way a related set of neural circuits (variants) can be configured or, for the case of neural events, incremental differences in variables chosen to represent those events. Each point within the domain bounded in blue represents a selector circuit (SC), defined as a neural circuit or neural event that acts to evoke a particular experiential quale (the **A**s in this case). Black points outside the domain represent circuit or event variants that fail to evoke that quale, or in this case, any experience at all. For an individual brain, the quale in question could theoretically be evoked by many SCs acting in concert, represented here by a point cloud, or few, or only one. Individual brains could thus vary both quantitatively (how many variants are active and effective) and qualitatively (how tightly clustered they are in SC-space). Experience **A*** is included to indicate that there may be differences in the experience evoked at different points within the domain, i.e., that experience **A*** may be qualitatively different than **A**. For example, the domain as a whole might include SCs for both fear (**A**) and panic (**A***), so a point cloud clustered predominantly on the left would evoke fear, on the right, panic, or both feelings together if the SCs are evenly distributed. How gradual the transition might be from fear to panic along trajectories in SC-space is not specified, nor how abruptly either experience is degraded for SCs located near the domain boundary, which could for that reason be "fuzzier" than shown here. The domain could also be more cloud-like in being diffuse and full of holes representing SCs inside the domain that happen not to produce an experience. Indeed, the term point cloud would typically be employed topologically to refer to the domain itself, so as to include the total set of all possible SCs of a specified type, but is here used in a more restricted sense, to refer to only those SCs realized in real brains at either the individual or population level. The diagram is highly schematic in reducing the high-dimensional space required to represent the complexities of real neural circuits and events to a two-dimensional surface, and is intended to apply only to the most fundamental units of experience, i.e., qualia.

and how they are to be analyzed. Topological methods are used elsewhere in the study of consciousness (e.g., Clark, 1996, 2000; Matthen, 2005; Rosenthal, 2010; Raffman, 2015), but not for the purpose of modeling evolutionary change.

CATEGORIZING CONTENTS

It is important first to distinguish contents of consciousness that are suitable for the analysis that follows from those that are not. To avoid any confusion, the term "contents" is not meant here to refer to anything more mysterious than a list written down on a piece of paper, and in no way implies that consciousness has the properties of a vessel that needs filling, or is limited in what it can contain. Though these both may be true, they are irrelevant to the analysis. The relevant point is that the contents of consciousness vary in complexity, from simple sensations, like the sharp pain from the prick of a needle or the feeling of pleasure, anxiety or fear, to the visual, auditory and cognitive experiences of such activities as hunting prey, avoiding predators, or comprehending a lecture on cognitive neuroscience. Since my concern here is with the elaboration of experience from simple beginnings, the analysis is restricted to those contents that might reasonably be supposed to have emerged early in evolution, and hence were available to be employed as components of later evolving, more complex contents. To this end I make following conjecture: that much as molecules are constructed of atoms, complex experiential contents are constructed of multiple elements among which are more fundamental units that are themselves contents, but are irreducible. So, to continue the analogy, molecules are reducible by chemical means while atoms are not, hence the most fundamental units of consciousness, whatever those are, will be those that involve no procedural sub-processes, and resist deconstruction by any means we currently have at hand, whether verbal argument, physical intervention or mathematical analysis. In consequence, they cannot be apprehended except by direct experience, which makes them essentially equivalent to qualia as usually defined (Tye, 1995, 2018). I use the term here despite its detractors (see Kanai and Tsuchiya, 2012 for a defense) because a quale simply "is" and so is ineffable, like the classic example of perceiving the color red, which exactly suits my requirements.

The idea that qualia are fundamental units of experience is widespread in consciousness studies,[1] but I treat them here as fundamental also for purposes of analysis and as objects of selection. Investigating consciousness from an evolutionary perspective has its own focus and agenda (Lacalli, 2021), and neither have been well served by existing theory. Addressing the question of what form consciousness took early in its evolutionary history is difficult to say the least, but is essential if we are ever to understand the link between consciousness as we experience it and the ancestral condition from which that consciousness derives. The current paper represents an attempt to do precisely that, but the methodology adopted is only directly applicable to a subset of experiences, namely those provisionally identifiable as qualia. For many theories of consciousness the focus is as much if not more so on complex contents, i.e., those combining qualia with other products of neural activity. Vision exemplifies this greater level of complexity, as the visual display, which allows the whole of the visual field to be perceived at once, has an intrinsic geometry and viewpoint that can be analyzed and understood in its own terms (Merker, 2007, 2013; Williford et al., 2018). One can then reasonably suppose that the properties of the display arise at least in part from the way visual input is processed and integrated, which will involve procedural rules, and so is sequential, algorithmic, and by analogy, computational (Wood, 2019). Hence the perception of a visual field, as an experience, is not a fundamental unit of consciousness as defined

[1]https://en.wikipedia.org/wiki/Qualia

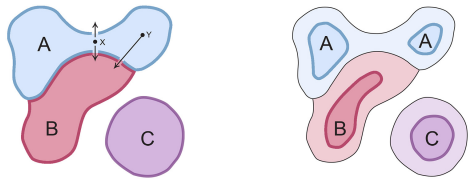


FIGURE 2 | How evolution acts to select the subset of SCs realized in real brains. Consider hypothetical domains A and B within whose boundaries the respective SCs evoke two distinguishable experiences, **A** and **B**. For SCs in real brains, and assuming SCs change location from generation to generation due to genetic and developmental variability, short trajectories that change or abolish an experience (e.g., X in the figure) are more likely to occur than longer ones (Y). Subjective experiences are therefore less robust in evolutionary terms when the SCs that evoke them are close to domain boundaries. Hence, over evolutionary time, the region within which point clouds are realized (bounded domains on the left hand diagram) will progressively shrink and separate from one another (right hand diagram) as the SCs in intervening regions of SC-space (paler colors) are eliminated from real brains. Domain C is included as a reminder that multiple domains can act together, as ensembles, so that, for example, experience **A** might only be evoked if both A and C (plus any number of additional domains) act in concert, or A and C might together evoke an entirely different experience.

above, and its dependence on neural circuits and patterns of activity make it too complex to be represented by a configuration space. Contents of this type, which are beyond the scope of this analysis, will be referred to as "formats." This would include vision, which, as a total experience, is a format. Similarly, memory dependence (Wilson and Sullivan, 2011) makes olfaction a format, though the NCCs responsible for evoking individual odors could potentially be mapped to a configuration space. Language would also be a format, for both its intrinsic structure and memory dependence (Chomsky, 1990; Jackendoff, 2002; Pinker and Jackendoff, 2005), as would everything that flows from the use of language, including reasoning, logic, and any form of conscious awareness with a linguistic component.

There are other ways of subdividing the contents of consciousness: between sensations and conscious thoughts (Block, 1995; Bayne and Montague, 2011), between phenomenal (P) and access (A) consciousness (Block, 1995), or core (CC) vs. extended (EE) phenomenal states in consciousness state space (Berkovich-Obana and Glicksohn, 2014), or through choosing a conservative vs. a liberal stance (Kemmerer, 2015). Most of these capture the distinction I've made above in one form or another, but for my purposes it is less important to determine where precisely the dividing line is drawn than to ensure that formats are excluded from consideration for being inherently too complex to map in a simple fashion. This avoids some of the conceptual difficulties highlighted by Velmans (2009), including the distinction between qualia and the reflexive or self-referential awareness of those qualia (Peters, 2014), and the "level" of consciousness is likewise not relevant (Overgaard and Overgaard, 2010; Bayne et al., 2016), as it might be affected by, say, sleep or anesthesia, so long as the qualia in question are unaltered in their character.

Treating qualia as more fundamental than more complex contents does not mean qualia necessarily evolved first. In fact the opposite would be the case if, as in many theories, the emergence of consciousness in evolution depended on algorithmic processes, e.g., of sensory processing, episodic memory or learning. The first content of consciousness would then have had at least some of the properties of format, but the sequence in which contents were added to evolving consciousness is not crucial to this analysis, nor is it a problem if there is some degree of dependence on algorithmic processes for most, if not all, conscious experience. Here I require only (1) that the set of all qualia, conceived of as fundamental units of experience, is not the null set, so that it is possible to have qualia that are not inextricably embedded in formats, and (2) that experiences that appear to be simple are indeed so, or at least can be dealt with as such, as qualia rather than formats. Three examples have then been selected that in my view provisionally pass muster in this respect: the simplest of tactile sensations, e.g., a sharp pain or itch (disregarding the means by which these are localized), the frequency range of sound, and the spectrum of light as we perceive it. These are used in the discussion of experience space in the section "Trajectories in Experience Space." To begin, however, it is necessary to consider how NCCs might be represented in a space that would map to experience space, where again, anything overtly format-like is excluded.

SELECTOR CIRCUIT SPACE: ROBUSTNESS AND ROUTES TO INNOVATION

There are multiple ways of constructing topological spaces to represent the physical factors that contribute to conscious experience: a space for mapping the genomic contribution to neural structure and activity, for example, or an NCC space mapping the neural correlates that underpin conscious experience, either as structural variables, activity-based variables or both. Because the genomic determinants of neural structure and activity are so far removed from the immediate mechanisms that evoke consciousness, the focus here is at the level of NCCs. The analysis could equally well be applied to any neural function, not just consciousness, excepting that, whereas there are various ways to model non-conscious neural circuits based on known examples, the absence of any consensus regarding what NCCs actually look like means that for consciousness, an indirect approach is currently the only available option.

A generally accepted definition of NCCs by Chalmers (2000) employs the idea of a mapping between the physical and the experiential: that NCCs are a "minimal neural system N such that there is a mapping of N to states of consciousness…" with caveats being that we need to be cognizant of whether N is both necessary and sufficient, or only the latter (Fink, 2016), and that correlates are not confused with markers or constituents of consciousness (Michel and Lau, 2020). Here I restrict the analysis to a subset of NCCs that I will refer to as selector circuits (SCs), defined as the neural circuits or activity patterns that serve as the proximate cause that a particular experience is evoked rather than some other. SCs would then fit into previously defined categories of core correlates (Block, 2005),

differentiating NCCs (Hohwy and Bayne, 2015), and difference makers of consciousness (DMCs, see Klein et al., 2020).[2] Klein et al. (2020) make the case for choosing difference makers over NCCs in the broad sense for their greater utility for dealing with causation in complex multi-component systems, the emphasis being on those correlates responsible for changing system output in a predictable way. Change is equally central to the conception developed by Neisser (2012) for isolating the causal component from an otherwise causally neutral set of neural correlates, a task that will be increasingly important as real data begin to emerge on NCC circuitry in real brains.

The kind of topological mapping I propose for SCs formally resembles one used by Fink (his figure 3, which maps "neural events"), but is more precisely defined, as a map of all possible configurations of those categories of circuits capable of acting as SCs, including variants that do not evoke an experience as well as those that do. The latter will then appear as islands, or domains, one for each experience, surrounded by a sea of variants with no selective effect on experience (**Figure 1**). SC-space is treated as a metric space, in which distance has an explicit meaning: distance between two adjacent points on the map will be defined as the smallest incremental change in the way SCs can be configured, or in the case of circuit activity, the smallest incremental change in the dynamical properties of the SCs in question. The cause of such changes could be genomic, e.g., due to mutation and recombination, or arise from variations introduced during brain development. I require here only that the incremental changes are observables of the system, available to a privileged observer to whom all physical features of the system are known, and are quantifiable, at least in principle. Thus, proximity in SC-space equates to similarity in neural structure or activity patterns, and proportionately greater distance reflects incrementally greater differences in these same variables. Expressing this in a two-dimensional map is clearly inadequate when even a moderately complex neural circuit will have myriad structural and activity-based features that can be configured in many different ways. The system then has many degrees of freedom that can only be fully captured in an n-dimensional space for very large n. Here, for purposes of illustration, $n = 2$ will suffice, with the caveat that there will be artifacts of this compression, e.g., that much of the incremental character of changes in higher dimensional space may be lost when mapped to one of lower dimension.

Consider next, with reference to **Figure 1**, how an SC for a given experience would be represented: as a single point in SC-space or a grouping of points. There would be a single point if, for an individual brain, the experience in question was evoked by either a single neural event or a set of exact, simultaneous replicates of that event. But if multiple events are required that exhibit some degree of variation, e.g., in the precise architecture of the circuits involved, the timing of events, or any other feature that makes them less than identical and simultaneous, the result is a point cloud. The position of the point (for a single event) or the point cloud (for multiple distinguishable events), and the degree of dispersion of the point cloud will, in the real world, vary between brains. In consequence, the experience evoked can potentially vary as well, so that a pinprick, for example, would be experienced differently from one individual to another, but each would still recognize the experience as painful. The key issue then, from an evolutionary perspective, is to determine which distribution of points in SC-space is most robustly buffered against being degraded over evolutionary time, that is, from generation to generation. The same question applies at the population level, where the SCs would necessarily map as a point cloud representing variation across the population. The consequences of occupying a less-than-optimal location in SC-space are different in these two cases, however. For an individual brain, a shift in position in SC-space will directly affect the experience, e.g., by enhancing, degrading or abolishing it. At the population level, this translates into an increased incidence of either enhanced or impaired experience across the population as a whole, and increased or reduced fitness for some individuals as compared with others.

Consider the case of an individual brain in more detail. We do not know how much mechanistic redundancy is built into the circuitry involved in sensory processing and consciousness (Hohwy and Bayne, 2015), but assuming there is some, the result in SC-space is a point cloud that, if highly localized, produces a combined experience that sums the separate contributions of component circuits that are nearly identical. For a more dispersed cloud there is a greater chance that the resultant experience combines components that are significantly different in character (e.g., that experience **A** in **Figure 1** might differ significantly from **A***). Having a larger and more disperse point cloud thus risks degrading the experience for an individual brain because some SCs will be altered to the point where they either make no contribution to the experience or introduce an element belonging to some distant variant of that experience. Assuming this is disadvantageous, selection will act to minimize the likelihood of it happening, giving localized point clouds an evolutionary advantage over larger diffuse ones. In consequence, the SCs produced over evolutionary time by real brains should map to a progressively shrinking subregion within their respective domains, at both the individual and population level, as they are extinguished from regions near domain boundaries (shown in pale colors in **Figure 2**). Redundancy is also a consideration. If there is little redundancy, meaning one or a few SC variants are required per brain to evoke an experience, then the reliability of the result depends on those few SCs being precisely replicated in each generation. With greater redundancy, meaning larger numbers of SCs, the deleterious effect of a few of these either degrading or otherwise altering the experience is reduced. Hence redundancy, coupled with stabilizing selection, will buffer the system against the maladaptive randomizing effects of mutation, recombination, and developmental variation as these impinge on individual SCs. Data on real SCs should also then show a positive correlation between the fraction of the potential domain to which those SCs map and the tightness of control exercised over their development.

[2]My choice of SCs over a more neutral term, such as selectors (Ss) or DMCs, in part reflects a mechanistic bias, but also makes the resulting configuration space easier to comprehend and explain.

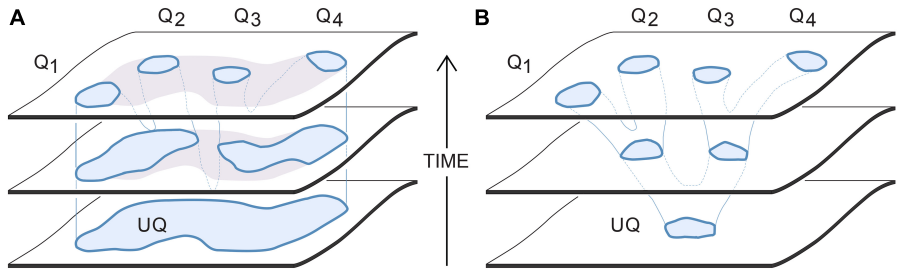


FIGURE 3 | Two options for how a set of domains on SC-space, representing the SCs that evoke specific qualia (the Qs), could derive by common descent from those evoking a single ur-quale (UQ). The intent is to show how the ur-quale is changed in character (horizontal axis) over evolutionary time (vertical axis). **(A)** The evaporating puddle scenario: this assumes an ur-quale whose SCs occupy a large domain, which is not then precisely defined at first with regard to the experiences those SCs evoke. A range of sensations would hence be evoked together from which the descendant qualia are progressively refined. Since the SCs remain within the parent domain, each newly evolved quale would incorporate elements present in the ur-quale. **(B)** The branched tree scenario: this assumes the SCs evoking the single ur-quale were distinct and well defined from the start, so the initial point cloud would have been restricted to a smaller domain compared with the puddle scenario. Since the branches of the tree diverge, all of the Qs in the tree scenario (in this example, all but Q3) will differ qualitatively from the ur-quale from which they all derive. See text for details.

Figure 2 illustrates the above arguments graphically using three domains (A, B, and C) representing regions in SC-space where SCs localized to A and B evoke, respectively, distinguishable experiences **A** and **B**. For a large domain, many different SC variants would map to the same experience. Whether large domain size is advantageous in and of itself, natural selection has no way of controlling this because domain size for a particular experience is an ontological given, belonging to the realm that Godfrey-Smith (2019), for example, refers to as "the physical." But what evolution can do is adjust the fraction of the domain that is occupied by the SCs of real brains. Whether the SCs act singly or in combination, what this means in practice is that SCs too near domain boundaries will be progressively eliminated, because small changes in map position alter the experience evoked (arrows from X in the left panel, which either abolish **A** or convert it to **B**) more easily than more distant points (arrow from Y), making the former less robust to genomic change and developmental variation. Assuming evolution favors robustness, the SC variants that survive selection will occupy a progressively smaller proportion of the original domain, so the point clouds of SCs formed by real brains both diverge and are reduced in size as shown in right panel.

But how would such domains arise in close proximity in the first place? Since only small changes in configuration are needed to alter the experience evoked, the underlying mechanism for evoking **A** and **B** would in such cases be similar, sharing many common features. The implication is that A and B are evolutionarily related, raising the possibility that they arose by common descent from an ancestral domain whose SCs once evoked an undifferentiated combination of **A** and **B** together. Refining this ancestral experience (an ur-quale in this formulation) so that **A** and **B** diverge, would have meant selecting brains where the activity of SCs mapping to A are increasingly correlated with each other, but not with those localized to B, and vice versa, and arranging for behavior to depend on this difference. By way of example, suppose one of the degrees of freedom represented by distance across SC-space relates to the timing of relevant neural events, e.g., either in frequency or duration. What we would then see is one set of frequencies or durations evoking **A** more than **B**, and eventually, by selection of variants, evoking **A** to the exclusion of **B**. By this means an initially large SC domain could, in principle, be repeatedly subdivided to produce a range of progressively more refined and precisely specified experiences.

Domain C is included in **Figure 2** as a reminder that there is a second route toward innovation, by addition and combinatorial action rather than subdivision. If we think of SC-space as defined so as to represent all possible SCs, evolution is, in effect, exploring a configuration space where any point in that space potentially represents a novel circuitry variant that would either alter an existing experience or evoke an entirely new one. Thus, C could evoke novel experience **C**, or A and C acting together might evoke that same **C**. Further, there could be any number of such distinct C-like domains, i.e., D, E, F, and so on, acting in combinatorial ways, and they need not be linked by descent. Encountering them allows evolution to expand the range of qualia that are experienced, while ensuring at the same time that they are robustly isolated from one another in terms of distance across SC-space. This is especially the case for new domains in distant parts of an n-dimensional SC-space, because the circuitry involved would then be well separated from other SCs by many configurational differences.

Of the various ways qualia might diverge from one another over evolutionary time, **Figure 3** shows two ends of a spectrum of possibilities, and can be interpreted as applying either at an individual or population level. However, at the individual level it is more meaningful (and this account will assume) that we are dealing with a situation of high redundancy, i.e., where multiple SC replicates act in concert. We can then have a situation, as in **Figure 3A**, where the ur-quale is evoked by a point cloud of SCs distributed over a large domain capable of evoking a multiplicity of qualitatively different sensations combined together in a single resultant experience. The sequence of progressive refinement would follow what I have chosen to call the evaporating puddle scenario ("puddle" for short), by analogy to the uneven evaporation of a large shallow puddle, leaving

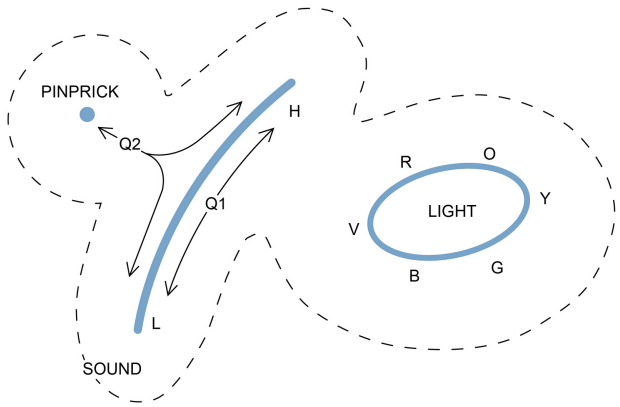


FIGURE 4 | A way of representing three experiences on a two-dimensional E-space. A pinprick is suggested here as a simple example of tactile experience, disregarding its localization, that can be represented as a point. The sensation of itch might be equally suitable. Sound, for animals that can distinguish frequencies, would be a line from low (L) to high (H) frequency. Color, as we experience it, is a closed curve, as the sequence from red to orange, yellow, green, blue, and violet (R, O, Y, G, B, V) is recursive, leaving the center of the curve for their blended combination, white light. The dotted lines are a reminder that, if these experiences are to be plotted together, there must be a zone of exclusion between them that is devoid of realized experience, as the three experiences would otherwise risk being combined in ways that would render them less distinguishable. The diagram could well have looked quite different if we consider the evolutionary past, how the three qualia originated, and the degree of homology between them. This is shown by the trajectories (arrows from Q1 and Q2). Trajectories originating at Q1 show a route by which a frequency-dependent acoustic experience might have evolved from an ur-quale that originally produced a much more limited range of that experience. Trajectories radiating from Q2 show routes by which an ancestral ur-quale common to multiple mechanosensor-based experiences might have evolved so as to separately evoke sound and tactile sensations, making these homologous as mechanosensations. The trajectories represent sequences of states that have changed over time, but points along the Q2 trajectories are ones that would have been present only in past brains, not present ones, as the SCs responsible for evoking intermediates between the qualia in question would long since have been extinguished by selection. For qualia unrelated through homology as experiences, there may be no such intervening points, and hence no access to intermediate experiences. This could be the case for light and sound for example, which share no obvious qualitative features, in which case there would be no justification for even trying to map them to the same surface. The reader is encouraged to think about how the figure might be used to illustrate the differences between a puddle-like evolutionary sequence and a tree-like one, i.e., to construct an E-space counterpart of **Figure 3**.

smaller residual puddles behind within the original outline. By analogy, in this scenario, as evolution progressively eliminates some SCs, those that remain would respond to sensory input by evoking a progressively more restricted set of qualia, each representing an element of experience present in the ur-quale from which they all derive. An example might be an ur-quale that, in this ancestral condition, combined together an assortment of negative feelings, such as fear, anxiety, panic, despair and disgust (see Panksepp, 2016) that come to be experienced separately by more highly evolved brains. The second alternative is the tree scenario (**Figure 3B**) where the SC variants are more tightly clustered from the outset, in a small domain, so as to produce an ancestral ur-quale of a more restricted kind. Over time, the original domain could then spawn sub-domains that diverge, like branches from a stem, so that the new experiences evoked by the SCs in each subdomain become realized contents. The experiences themselves are then well defined throughout, but change incrementally in character as evolution explores surrounding regions of SC space. Because the SC point cloud is small from the start, a higher degree of developmental precision would be required throughout this branching process compared with the puddle scenario. Also, since the tree fans outward over time, novel, divergent experiences can evolve that differ in significant ways from the ur-quale.

One can then ask, of all the qualia we experience, how many, if any, trace their origins to patterns of the above kind, and hence are related through homology. A plausible conjecture is that this is most likely to be the case for qualia sharing related sensory modalities. Obvious examples would be sets of related emotional states, e.g., the negative feelings referred to above, the different acoustic tones we hear, or the spectral colors that arise in vision. One can also ask, since SC-space is a configuration space rather than a real space, if this analysis provides any clues about the number of neurons or volume of tissue required to implement a set of SCs. The answer is that it does not, because the physical volume occupied by the configuration representing a given point in SC-space, whether large or small, is not specified. Consequently this account makes no claims about the actual size, structure or complexity of SCs, and includes no circuitry diagrams, because there is no way currently to choose between many possible options. SCs could be subcomponents of large diffuse cortical networks, or small localized circuits of a few neurons; they could depend on structural features such as the way active synapses are deployed in 3D space, or on activity patterns where it is the pattern itself that exerts a selective action. What can be said is that redundancy matters, and if multiple SCs of similar type must act in concert, implementing this should require a greater volume of tissue than if there is no such redundancy.

Finally, recall that for real populations, there is the problem of maintaining an optimal set of SCs from generation to generation against the degrading effects of random genomic and developmental events. It is a matter of conjecture how rapidly, in the absence of selection, this would happen, but there is no reason that the rate should be the same for both simple and complex contents, i.e., for qualia as compared to the more complex experiences I have here categorized as formats. For qualia, the issue is how reliably some SC variants are formed rather than others. In contrast, for formats, robustness depends on the reliability of reproducing, in each generation, the circuits that execute the algorithms on which each format depends, which are almost certainly different from, and independent of the SCs responsible for evoking the qualia themselves. Hence there is a real possibility that formats can be more robust than the qualia they employ. This could have practical consequences where formats have come to dominate behavioral decision-making, as they have for our own species. In this sense, the distinction made here between qualia and formats is important, not only as a theoretical construct, but for its possible real-world implications.

TRAJECTORIES IN EXPERIENCE SPACE

The SC-space considered above is a way to represent the measurable properties of a real physical system, i.e., of circuits and their activity. E-space is different in attempting to represent the non-physical properties of experience itself. It is not then a configuration space in the usual sense, as there is nothing physical to configure, but it is a metric space where distance has a specific meaning: that each increment in distance is the minimal distinguishable difference between two experiences. To avoid complicating this definition with issues of a strictly subjective nature, such as whether one can be sure that two different tones of sound are as distinguishable from each other as either is from a flash of light, I will add the further criterion that any two adjacent points are those between which no experience can be inserted that is not intermediate between the two. So, for example, the only experience between two acoustic tones would be another acoustic tone, which disallows a flash of light or a noxious odor from occupying that location. In practical terms, this means adjacent points will belong to sensory experiences that are either the same or only incrementally different. This raises the point of whether different qualia in fact share features that allow them to be mapped together on the same surface, a question which, as discussed below, may depend on whether they are related through homology.

Other topological constructs have been used to investigate conscious experience. Of these, E-space as defined here differs from quality space, which maps subjective experience quantitatively, and which from my perspective is problematic when applied to complex sensory states, such as vision (see critique by Matthen, 2004), but also in other applications (Kostic, 2012; Young et al., 2014). I likewise distinguish between E-space and similarity space, which maps experiences with respect to their similarities and differences, as my concern is less with the relation between physical stimuli and the sensations they evoke than with how, in principle, neural circuitry acts to shape subjective experience.

To this end, E-space is treated here as the space of all possible qualia regardless of whether they are experienced by any particular brain. This is meant simply as a convenience for this particular thought experiment, not to argue in support of the view that the ultimate source of conscious experience lies outside the biological realm. It also means that E-space will be larger than the subdomain available to a given brain, so that evolution can be thought of as acquiring novel qualia as it explores E-space through neural innovation. The human brain, for example, might have the potential for an experience equivalent to a bat's, during echolocation, by evoking it from regions of E-space that are available to human brains, but have been rendered silent by evolution. Or, it may be that human brains have never had access to those regions of E-space. There could also be many experiences that no vertebrate brain has yet evolved to evoke, but what these might be is, from a human perspective, impossible to judge.

The properties of E-space defined in this way can be illustrated with three examples that map as a point, a line and a closed curve (**Figure 4**). The first is how I would represent pinprick, which so far as I can see (or, literally, feel), is an experience so simple that, stripped of positional reference, it lacks any other aspect; it simply "is." For the second, a line, I have chosen the range of tonal sound as registered by the cochlea, where the auditory experience varies in a graded way depending on vibrational frequency, but terminates at some point at both ends. For my purposes it does not matter whether each tone is treated as a distinct quale, or whether sound at different frequencies is a single quale that is "tunable" in some way. What matters is that all other qualia are excluded from the line of tonal experience because they are not intermediate between any two tones. My third example is the visual perception of color, where there is a continuous gradation in the nature of the experience, but no point of termination because the colors, at least as we experience them, form a continuous and recursive sequence (Matthen, 2005, 2020). Combining colors moves you toward the center of the curve, where the color is replaced by white light, so trajectories across the domain enclosed by the curve are graded as required.

The figure shows all three qualia together on the same two-dimensional plane as a way of illustrating a feature that is necessary regardless of how many dimensions the map is intended to represent, and that is divergence. That is, if all three are mapped together, and for the way the metric is defined, the three will be separated by a zone of exclusion surrounding each one (inside the dashed lines in the figure) because qualia too similar to one another risk being indistinguishable in practice, especially at low intensity, which makes them maladaptive from an evolutionary standpoint. It is, after all, at the margins of perception that selection will often exert its strongest effect, e.g., that the antelope that is only slightly less able than other herd members to distinguish between different sensory cues is the one that gets eaten.

The question then is, under what conditions is it appropriate to map diverse qualia to the same topological space. This is ultimately a question about the nature of qualia themselves: are they comparable in kind in the sense that they could in principle grade into one another, or not? With clearly related qualia, such as a set of acoustic tones, one can suppose this is the case, i.e., that they are both similar in character and grade into one another in an a continuum. And it is plausible that they may share a common origin, as an acoustic ur-quale, represented by Q1 in **Figure 4**. Indeed all experiences of mechanosensory origin (touch, pressure, vibration, hearing) could conceivably derive from a common ur-quale, positioned like Q2 in **Figure 4**. From this point, the incremental divergence required to evolve the experiences of pinprick and hearing would define a surface by tracing out a trajectory of points in E-space that do in fact exist, because they have existed in the past in real brains. That part of the surface is hence a valid construct in reality. In contrast, considering the qualitative difference between the experiences of light and sound, with no obvious intermediate between them, there is no reason to suppose they could be mapped together. This is reinforced by what we know of the sensory cells involved, that they have evolved from separate receptor-based systems (Schlosser, 2018), so homology between light and sound as experiences is possible, but not expected. Assessing homology can be problematic, however (Hall and Kerney, 2012), the complication here being

that judging whether two experiences are homologous based on common descent is quite separate from the issue of homology as it relates to the underlying neural circuits, and these circuits will almost certainly share many common features irrespective of whether the qualia they evoke are homologous at the experiential level.

The advantage of dealing with qualia that are potentially homologous is that a more plausible case can be made for an isomorphic mapping between SC-space and E-space. That is, where patterns of past divergence follow a tree or puddle pattern, E-space might exhibit a matching pattern of diversification and divergence. For sound, for example, the range of frequencies experienced might, in SC-space, be evoked by a continuous sequence of SC domains that map in an orderly fashion to a corresponding line in E-space. This would have the advantage of being a parsimonious explanation, the problem being that we do not know if anything concerning consciousness is, in fact, parsimonious. Evoking new sound experiences across a frequency range might instead depend on the addition of multiple new domains scattered all over SC-space acting in combinatorial ways. Further, distances need not map proportionately, since a short displacement in SC-space could yield a large one in E-space, while a large displacement in SC-space might make no difference at all to the experience. The conclusion is that for qualia sharing common descent, it is possible that there could be an isomorphic mapping between SC- and E-space, but this is by no means the only option.

To conclude this section, it is useful to make a remark on referral, sometimes included among the hard problems of consciousness, e.g., by Feinberg (2012). Take vision, for example, considered here as a format, where the inherent viewpoint ensures that the experience is perceived as external, i.e., it is referred to the outside world (Merker, 2013). The provisional conclusion one might then draw is that referral is a property of any format structured so as to ensure this result, and that other mappings, including the somatosensory map, would share this property. But this is not the only possibility. Consider instead a somatosensory experience that was more akin to the acoustic experience of different frequencies. The conscious sensation of touch at different points along the rostro-caudal axis of the body would then be distinguishable in the same way as acoustic tones generated by the stimulation of different hair cells along the axis of the cochlea. The position-specific aspect of the somatosensory experience would thus be due to a graded or tunable quale, but to a single quale none the less, rather than a format. I mention this as a possibility, not so much to argue the case, but to illustrate the fact that we cannot predict in advance, or even judge from our own experience, the limits of what evolution is capable of doing with the qualia at its disposal.

CONCLUSIONS, AND THE FUNCTION OF CONSCIOUSNESS

This account proposes a conceptual framework, using a configuration space analysis, for investigating how evolution acts on the selector circuits (SCs, a subset of NCCs) responsible for evoking a particular conscious content as opposed to any other. The analysis depends on the supposition that there are fundamental units of experiences (qualia) that are distinguishable from more complex contents of consciousness (here, formats), and that qualia can be dealt with individually both at an analytical level and as objects of selection. But there are two further considerations. First, a caveat, that there is good reason to doubt that all contents will yield to the same set of analytical methods, and in particular, that a configuration space applicable to qualia can be usefully applied to formats. And second, a result of the analysis, that the question of evolutionary descent is a significant one, in that qualia that are homologous as experiences are intrinsically more easily dealt with in relation to one another than those that are not. This has practical implications for a future where we have more access to real data on NCCs relevant to various forms of experience, the expectation being that SC-type NCCs will exhibit both constant and variable features, but the variability will be least between qualia sharing common descent.

There is a developmental aspect here as well, since it is the variability among developing brain circuits, and the synaptic plasticity on which this variability depends, that provide the raw material for evolutionary innovation. For consciousness, and for SCs in particular, there are mechanisms that would allow this variability to be harnessed so as to ensure a precisely controlled outcome (Lacalli, 2020). Variability in this case means that the synaptic networks in question can be dynamically reconfigured as they develop, which means a degree of synaptic plasticity is an inherent part of the process. Synaptic plasticity is most frequently dealt with in relation to its role in real-time cortical functions like learning and memory (e.g., Attardo et al., 2015), but for SCs, in contrast, plasticity must diminish at some point during development if the resulting structure is to be stable in real time, and hence produce conscious experiences that are themselves stable. There should consequently be a division of labor among neural circuits, such that those involved in functions requiring real-time plasticity on a continuing basis, like memory, are precluded from involvement in those aspects of consciousness requiring real-time stability, including the evocation of qualia. This has implications for how the different functions associated with the production of conscious sensations are distributed across the brain and its various substructures.

As a final point, the configuration space representation can be used to illustrate something quite precise about the function of consciousness from an evolutionary perspective. I have expressed this previously as follows (Lacalli, 2020, p. 6): that consciousness functions as "a mechanism for restructuring synaptic networks in ways that would not otherwise have occurred, in order to produce advantageous behavioral outcomes that would not otherwise have happened." Topologically, this is saying that there are regions of SC-space, and hence E-space, that cannot in practice (i.e., in real brains) be accessed except through the agency of natural selection acting on the outcome of consciously controlled behaviors. A consideration of SC-space shows why: that for every

region in SC-space that evokes a particular conscious experience, there is a boundary a finite distance away that separates points within that domain from those outside it (cf. **Figure 1**). Starting from outside the domain, it is possible in principle for a fortuitous change to the genome to move the system in one jump from that starting point to deep within the domain. Hence a specific and reliably evoked quale could theoretically emerge from the non-conscious condition at one jump. But the spatial metric used here means that moving "to deep within" the domain would require multiple changes in the genome, or one change with multiple consequences for development of a very precise type, which means that the chance of this happening randomly is vanishingly small. Evolution achieves this instead through natural selection acting at a population level over multiple generations because, and only because, consciousness has an adaptive advantage over the absence of consciousness at each generational step. Hence, the function of consciousness from an evolutionary perspective is to provide access to otherwise inaccessible points in SC-space (indeed, in NCC-space more generally) and, correspondingly, in E-space. This may seem an unsatisfying conclusion, because it tells us nothing about the proximate purpose for which consciousness evolved, but it is the more general answer, and hence conceptually the more meaningful one.

DATA AVAILABILITY STATEMENT

There is no data beyond that included in the article; further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

TL was solely responsible for the preparation and content of this article.

FUNDING

Funds to support this work were received from the L. G. Harrison Research Trust.

ACKNOWLEDGMENTS

I thank Björn Merker for a stimulating exchange of ideas on these subjects, the reviewers for their helpful comments, and Riley Lacalli for preparing the figures.

REFERENCES

Abbott, A. (2020). What animals really think. *Nature* 584, 183–185.

Attardo, A., Fitzgerald, J. E., and Schnitzer, M. J. (2015). Impermanence of dendritic spines in live adult CA1 hippocampus. *Nature* 523, 592–596.

Baron, A. B., and Klein, C. (2016). What insects can tell us about the origins of consciousness. *Proc. Nat. Acad. Sci. U.S.A.* 113, 4900–4908.

Bayne, T., and Montague, M. (2011). "Cognitive phenomenology: an introduction," in *Cognitive Phenomenology*, eds T. Bayne and M. Montague (Oxford: Oxford University Press), 1–34.

Bayne, T., Hohwy, J., and Owen, A. M. (2016). Are there levels of consciousness? *Trends Cogn. Sci.* 20, 405–413. doi: 10.1016/j.tics.2016.03.009

Berkovich-Obana, A., and Glicksohn, J. (2014). The consciousness state-space (CSS)–a unifying model for consciousness and self. *Front. Psychol.* 5:341. doi: 10.3389/fpsyg.2014.00341

Black, D. (2021). Analyzing the etiological functions of consciousness. *Phenom. Cogn. Sci.* 20, 191–216. doi: 10.1007/s11097-020-09693-z

Block, N. (1995). On a confusion about a function of consciousness. *Behav. Br. Sci.* 18, 227–287. doi: 10.1017/S0140525X00038188

Block, N. (2005). Two neural correlates of consciousness. *Trends Cogn. Sci.* 9, 46–52. doi: 10.1016/j.tics.2004.12.006

Brown, R., Lau, H., and LeDoux, J. E. (2019). Understanding the higher-order approach to consciousness. *Trends Cogn. Sci.* 23, 754–768. doi: 10.1016/j.tics.2019.06.009

Butler, A. (2008). Evolution of the thalamus: a morphological and functional review. *Thal. Rel. Syst.* 4, 35–58. doi: 10.1017/S1472928808000356

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *J. Cons. Stud.* 2, 200–219. doi: 10.1093/acprof:oso/9780195311105.003.0001

Chalmers, D. J. (2000). "What is a neural correlate of consciousness?" in *Neural Correlates of Consciousness: Empirical and Conceptual Problems*, ed. T. Metzinger (Cambridge, MA: MIT Press), 12–40. doi: 10.1093/acprof:oso/9780195311105.003.0003

Chomsky, N. (1990). "On the nature, acquisition and use of language," in *Mind and Cognition: A Reader*, ed. W. G. Lycan (London: Blackwells), 627–645.

Clark, A. (1996). *Sensory Qualities*. Oxford: Oxford University Press, doi: 10.1093/acprof:oso/9780198236801.001.0001

Clark, A. (2000). *A Theory of Sentience*. Oxford: Oxford University Press, doi: 10.1093/acprof:oso/9780198238515.001.0001

Dainton, B. (2000). *Streams of Consciousness: Unity and Continuity of Conscious Experience*. London: Routledge.

Dehaene, S., and Naccache, C. (2001). Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition* 79, 1–37. doi: 10.1016/S0010-0277(00)00123-2

deVries, J., and Ward, L. M. (2016). An "ecological" action-based synthesis. *Behav. Br. Sci.* 39:e173. doi: 10.1017/S0140525X15002046

Feinberg, T. E. (2012). Neuroontology, neurobiological naturalism, and consciousness: a challenge to scientific reduction and solution. *Phys. Life Revs.* 9, 13–34. doi: 10.1016/j.plrev.2011.10.019

Feinberg, T. E., and Mallatt, J. M. (2016). *The Ancient Origins of Consciousness*. Cambridge, MA: MIT Press.

Fink, S. B. (2016). A deeper look at "neural correlates of consciousness". *Front. Psychol.* 7:1044. doi: 10.3389/fpsyg.2016.01044

Godfrey-Smith, P. (2016). "Animal evolution and the origins of experience," in *How Biology Shapes Philosophy: New Foundations for Naturalism*, ed. D. L. Smith (Cambridge, UK: Cambridge University Press), 51–71. doi: 10.1017/9781107295490

Godfrey-Smith, P. (2019). Evolving across the explanatory gap. *Philos. Theor. Pract. Biol.* 11:1. doi: 10.3998/ptpbio.16039257.0011.001

Hall, B., and Kerney, R. R. (2012). Levels of biological organization and the origin of novelty. *J. Exp. Zool. B (Mol. Dev. Evol.)* 318, 428–437. doi: 10.1002/jez.b.21425

Hohwy, J., and Bayne, T. (2015). "The neural correlates of consciousness: causes, confounds and constituents," in *The Constitution of Phenomenal Consciousness: Towards a Science and Theory*, ed. S. M. Miller (Amsterdam, NL: John Benjamins Publ. Co.), 155–176. doi: 10.1075/aicr.92.06hoh

Jackendoff, R. S. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.

Kanai, R., and Tsuchiya, N. (2012). Qualia. *Curr. Biol.* 22, R392–R396. doi: 10.1016/j.cub.2012.03.033

Kemmerer, D. (2015). Are we ever aware of concepts? A critical question for the global neuronal workspace, integrated information, and attended intermediate-level representation theories of consciousness. *Neurosci. Consious.* 2015, 1–10. doi: 10.1093/nc/niv006

Klein, C., Hohwy, J., and Bayne, T. (2020). Explanation in the science of consciousness: from neural correlates of consciousness (NCCs) to the difference makers of consciousness (DMCs). *Phil. Mind Sci.* 1:4. doi: 10.33735/phimisci.2020.II.60

Kostic, D. (2012). The vagueness constraint and the quality space for pain. *Phil. Psychol.* 25, 929–939. doi: 10.1080/09515089.2011.633696

Kostic, D. (2017). Explanatory perspectivalism: limiting the scope of the hard problems of consciousness. *Topoi* 36, 119–125. doi: 10.1007/s11245-014-9262-7

Lacalli, T. C. (2018). Amphioxus neurocircuits, enhanced arousal, and the origin of vertebrate consciousness. *Cons. Cogn.* 62, 127–134. doi: 10.1016/j.concog.2018.03.006

Lacalli, T. C. (2020). Evolving consciousness: insights from Turing, and the shaping of experience. *Front. Behav. Neurosci.* 14:598561. doi: 10.3389/fnbeh.2020.598561

Lacalli, T. C. (2021). An evolutionary perspective on chordate brain organization and function: insights from amphioxus, and the problem of sentience. *Philos. Trans. R. Soc. Lond. B*. doi: 10.1098/rstb.2020.0520

Levine, J. (1983). Materialism and qualia: the explanatory gap. *Pac. Phil. Quart.* 64, 354–361. doi: 10.1111/j.1468-0014.1983.tb00201.x

Levine, J. (2009). "The explanatory gap," in *The Oxford Handbook of Philosophy of Mind*, eds A. Beckman, B. P. McLaughlin, and S. Walter (Oxford: Oxford University Press), doi: 10.1093.oxfordhb/9780199262618.003.0017

Marques, J. C., Schaak, D., Robson, D. N., and Li, J. M. (2019). Internal state dynamics shape brainwide activity and foraging behaviour. *Nature* 577, 239–243. doi: 10.1038/s41586-019-1858-z

Matthen, M. (2004). Features, places, and things: reflections on Austen Clark's theory of sentience. *Phil. Psychol.* 17, 497–518. doi: 10.1080/09515080420003041 99

Matthen, M. (2005). *Seeing, Doing, and Knowing: A Philosophical Theory of Sense Perception*. Oxford: Clarendon Press.

Matthen, M. (2020). "Unique hues and colour experience," in *The Routledge Handbook of the Philosophy of Colour*, eds D. H. Brown and F. Macpherson (Oxford: Oxford University Press).

Merker, B. (2005). The liabilities of mobility: a selection pressure for the transition to consciousness in animal evolution. *Cons. Cogn.* 14, 89–114. doi: 10.1016/S1053-8100(03)00002-3

Merker, B. (2007). Consciousness without a cerebral cortex: a challenge for neuroscience and medicine. *Behav. Brain Sci.* 30, 63–81; discussion 81–134. doi: 10.1016/B978-044452977-0/50010-3

Merker, B. (2013). The efference cascade, consciousness, and its self: naturalizing the first person pivot of action control. *Front. Psych.* 4:501. doi: 10.3389/fpsyg.2013.00501

Merrick, C., Godwin, C. A., Geisler, M. W., and Morsella, E. E. (2014). The olfactory system as the gateway to the neural correlates of consciousness. *Front. Psychol.* 4:1011. doi: 10.3389/fpsgy.2013.01011

Michel, M., and Lau, H. (2020). On the dangers of conflating strong and weak versions of a theory of consciousness. *Phil. Mind Sci.* 1:8. doi: 10.33735/phimisci.2020.II.54

Neisser, J. (2012). Neural correlates of consciousness reconsidered. *Cons. Cogn.* 21, 681–690. doi: 10.106/j.concog/2011.03.012

Oizumi, M., Albantakis, L., and Tononi, G. (2014). From the phenomenology to the mechanisms of consciousness: integrated information theory 3.0. *PLoS Comp. Biol.* 10:e1003588. doi: 10.1371/journal.pcbi.1003588

Overgaard, M., and Overgaard, R. (2010). Neural correlates of contents and levels of consciousness. *Front. Psych.* 1:164. doi: 10.3389/fpsyg.2010.00164

Panksepp, J. (2016). The cross-mammalian neurophenomenology of primal emotional affects: from animal feelings to human therapeutics. *J. Comp. Neurol.* 524, 1624–1635. doi: 10.1002/cne.23969

Peters, F. (2014). Consciousness should not be confused with qualia. *Logos Epist.* 5, 63–91. doi: 10.5840/logos-episteme20145123

Piccinini, G., and Bahar, S. (2013). Neural computation and the computational theory of cognition. *Cogn. Sci.* 34, 453–488. doi: 10.11111/cogs.12012

Pinker, S., and Jackendoff, R. S. (2005). The faculty of language: what's special about it? *Cognition* 95, 201–236. doi: 10.1016/j.cognition.2004.08.004

Raffman, D. (2015). "Similarity spaces," in *The Oxford Handbook of Philosophy of Perception*, ed. M. Matthen (Oxford: Oxford University Press), doi: 10.1093/oxfordhb/9780199600472.13.030

Rosenthal, D. (2010). How to think about mental qualities. *Phil. Issues: Philosophy of Mind* 20, 368–393. doi: 10.1111/j-1533-6077.2010.00190.x

Schlosser, G. (2018). A short history of nearly every vertebrate sense–the evolutionary history of vertebrate sensory cell types. *Integr. Comp. Biol.* 58, 301–316. doi: 10.1093/icb/icy024

Schmidhuber, J. (2015). Deep learning in neural networks: an overview. *Neur. Netw.* 61, 85–117. doi: 10.1016/j.neunet.2014.09.003

Shepherd, G. M. (2007). Perspectives on olfactory processing, conscious perception, and orbitofrontal complex. *Ann. N.Y. Acad. Sci.* 1121, 87–101.

Solé, R., and Valverde, S. (2020). Evolving complexity: how tinkering shapes cells, software and ecological networks. *Phil. Trans. R. Soc. Lond. B* 375:201190325. doi: 10.1098/rstb.2019.0325

Tosches, M. A. (2017). Developmental and genetic mechanisms of neural circuit evolution. *Dev. Biol.* 431, 16–25. doi: 10.1016/j.ydbio.2017.06.016

Tye, M. (1995). *Ten Problems of Consciousness*. Cambridge, MA: MIT Press.

Tye, M. (2003). *Consciousness and Persons: Unity and Identity*. Cambridge, MA: MIT Press.

Tye, M. (2018). "Qualia," in *The Stanford Encyclopedia of Philosophy*, ed. E. N. Zalta (Stanford, CA: Metaphysics Research Lab, Stanford University).

Van Gulick, R. (2018). "Consciousness," in *The Stanford Encyclopedia of Philosophy*, ed. E. N. Zalta (Stanford, CA: Metaphysics Research Lab, Stanford University).

Velmans, M. (2009). How to define consciousnessand how not to define consciousness. *J. Cons. Stud.* 16, 139–156.

Velmans, M. (2012). The evolution of consciousness. *Contemp. Soc. Sci.* 7, 117–138. doi: 10.1080/21582041.2012.692099

Ward, L. M. (2011). The thalamic dynamic core theory of conscious experience. *Cons. Cogn.* 20, 464–486. doi: 10.1016/j.concog.2011.01.007

Williford, K., Bennequin, D., Friston, K., and Radrauf, D. (2018). The projective consciousness model and phenomenal selfhood. *Front. Psychol.* 9:2571. doi: 10.3389/fpsyg.2018.0271

Wilson, D. A., and Sullivan, R. M. (2011). Cortical processing of odor objects. *Neuron* 72, 506–519. doi: 10.1016/j.neuron.2011.10.027

Wood, C. C. (2019). The computational stance in biology. *Phil. Trans. R. Soc. Lond. B* 374:20180380. doi: 10.1098/rstb.2018.0380

Woodruff, M. L. (2017). Consciousness in teleosts: there is something it feels like to be a fish. *Animal Sent.* 2:13. doi: 10.5129/2377-7478.1198

Young, B. D., Keller, A., and Rosenthal, D. (2014). Quality-space theory in olfaction. *Front. Psychol.* 5:1. doi: 10.3389/fpsyg.2014.00001

Yu, Y., Si, X., Hu, C., and Zhang, J. (2019). A review of recurrent neural networks: LSTM cells and network architectures. *Neural Comput.* 31, 1235–1270. doi: 10.1162/neco_a_01199

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Systems Neuroscience

HYPOTHESIS AND THEORY
published: 05 November 2021
doi: 10.3389/fnsys.2021.768762

Balancing Prediction and Surprise: A Role for Active Sleep at the Dawn of Consciousness?

*Matthew N. Van De Poll and Bruno van Swinderen**

Queensland Brain Institute, The University of Queensland, Brisbane, QLD, Australia

The brain is a prediction machine. Yet the world is never entirely predictable, for any animal. Unexpected events are surprising, and this typically evokes prediction error signatures in mammalian brains. In humans such mismatched expectations are often associated with an emotional response as well, and emotional dysregulation can lead to cognitive disorders such as depression or schizophrenia. Emotional responses are understood to be important for memory consolidation, suggesting that positive or negative 'valence' cues more generally constitute an ancient mechanism designed to potently refine and generalize internal models of the world and thereby minimize prediction errors. On the other hand, abolishing error detection and surprise entirely (as could happen by generalization or habituation) is probably maladaptive, as this might undermine the very mechanism that brains use to become better prediction machines. This paradoxical view of brain function as an ongoing balance between prediction and surprise suggests a compelling approach to study and understand the evolution of consciousness in animals. In particular, this view may provide insight into the function and evolution of 'active' sleep. Here, we propose that active sleep – when animals are behaviorally asleep but their brain seems awake – is widespread beyond mammals and birds, and may have evolved as a mechanism for optimizing predictive processing in motile creatures confronted with constantly changing environments. To explore our hypothesis, we progress from humans to invertebrates, investigating how a potential role for rapid eye movement (REM) sleep in emotional regulation in humans could be re-examined as a conserved sleep function that co-evolved alongside selective attention to maintain an adaptive balance between prediction and surprise. This view of active sleep has some interesting implications for the evolution of subjective awareness and consciousness in animals.

Keywords: REM sleep, consciousness, predictive coding, emotions, invertebrate

OPEN ACCESS

Edited by:
Eva Jablonka,
Tel Aviv University, Israel

Reviewed by:
Karl Friston,
University College London, United Kingdom
Yu Hayashi,
University of Tsukuba, Japan

***Correspondence:**
Bruno van Swinderen
b.vanswinderen@uq.edu.au

Received: *01 September 2021*
Accepted: *08 October 2021*
Published: *05 November 2021*

Citation:
Van De Poll MN and van Swinderen B (2021) Balancing Prediction and Surprise: A Role for Active Sleep at the Dawn of Consciousness? Front. Syst. Neurosci. 15:768762. doi: 10.3389/fnsys.2021.768762

INTRODUCTION

Why do we dream? Every human since the dawn of humanity must have asked themselves this bewildering question, which seems inextricably linked to another related question: why do we sleep? It is therefore quite astounding to note that it was only about 100 years ago that a distinct sleep stage was identified – rapid eye movement (REM) sleep – that seemed to be associated with vivid dream

reports (Loomis et al., 1937; Aserinsky and Kleitman, 1953), and that was different from other sleep stages such as slow-wave sleep (SWS; Blake and Gerard, 1937). Humans were probably always aware that other humans, or their animal companions, were engaging in different kinds of sleep. Their bed partners might twitch during their sleep sometimes or breathe deeply other times, their babies might suddenly smile, their dogs whined or padded the air with their paws (but only sometimes). These were all clues that different kinds of sleep were potentially at play, but it required the advent of brain recordings and electro-encephalography (EEG) in the last century to conclusively show, in humans as well as other mammals, that these were indeed distinct sleep stages. We now know that REM sleep is associated with wake-like electrical activity across the mammalian brain cortex, characterized by low-amplitude, desynchronized field potentials (Aserinsky and Kleitman, 1955; Jouvet, 1961; Hobson, 2009a). In contrast, with its unique high-amplitude slow waves (1–4 Hz 'delta' waves), SWS seemed different enough to wakefulness to have traditionally attracted more interest as somehow being 'real' or 'deep' sleep, potentially achieving some more crucial functions than REM sleep. Early on it was discovered that this distinct sleep stage, REM, was strongly associated with the subjective state of disconnected consciousness we term dreams, the often absurd or embarrassing nature of which did little to improve the standing of REM.

To date, almost every animal that has been investigated carefully (meaning, satisfying key behavioral criteria such as quiescence, increased arousal thresholds, and homeostatic regulation (Campbell and Tobler, 1984), has been found to need sleep. Beyond mammals and birds, this ranges from animals without central nervous systems (or 'brains') such as hydra (Kanaya et al., 2020) and jellyfish (Seymour et al., 2004; Nath et al., 2017), and roundworms (Raizen et al., 2008) to insects (Tobler and Neuner-Jehle, 1992; Shaw et al., 2000), fish (Zhdanova et al., 2001; Prober et al., 2006; Yokogawa et al., 2007), amphibians (Libourel and Herrel, 2016), and reptiles (Tauber et al., 1966; Ayala-Guerrero and Mexicano, 2008). All these animals become periodically quiescent (i.e., immobile) in order to engage important biological processes that are largely incompatible with waking activity and ongoing behavior. These processes include cell repair mechanisms, growth and development, waste and metabolite clearance, and stress regulation (Sassin et al., 1969; Xie et al., 2013; Ogawa and Otani, 2014; Tononi and Cirelli, 2014). In humans and other mammals, these basic cellular sleep functions typically occur during SWS, when the cortex is traversed by slow 'delta' waves (Dijk et al., 1990) but the rest of the brain is more quiet (Siegel, 2008). This suggests that ancient sleep functions important for maintaining neuronal health have been packaged into SWS in mammals and birds, and that the slow (1–4 Hz) waves characteristic of SWS in these animals are probably a thalamocortical novelty riding on a more ancient drive for periodic brain quiescence. All animals appear to need such periodic neural quiescence in order to develop and adapt appropriately to their environment. In contrast, only a subset of animals seem to engage in REM sleep (**Figure 1**).

During REM sleep, the brain looks awake but animals remain significantly less responsive to the outside world (Green and Arduini, 1954), so based on increased arousal thresholds alone this has qualified as 'sleep' (Andrillon and Kouider, 2020). Since this is potentially confusing (why are we then not awake and responsive?), REM sleep has also been termed 'paradoxical sleep' (Jouvet-Mounier et al., 1969) or 'active sleep' (Libourel and Herrel, 2016). The recent discovery of REM sleep-like sleep in disparate animals such as reptiles (Shein-Idelson et al., 2016), fish (Leung et al., 2019), and molluscs (Iglesias et al., 2019; Medeiros et al., 2021) casts doubt on a common evolutionary origin for REM sleep and instead suggests a selective pressure to achieve related 'active sleep' functions in these diverse creatures. What might these functions be? While 'deep sleep' functions seem easier to comprehend (i.e., recurrent neural quiescence is required for achieving cellular homeostasis), why should some animal brains remain wake-like but disconnected from the outside world? This seems a potentially hazardous prospect, with some cuttlefish for example engaging in striking chromatophore pattern displays during this purported sleep stage (Frank et al., 2012; Iglesias et al., 2019) – clearly not a good idea for an animal not paying attention to potential predators. REM sleep must therefore be performing an important function (or multiple functions), to offset the disadvantage of being disengaged from the immediate environment. That active as well as deep sleep stages might even be required for the smallest animal brains, such as flies (van Alphen et al., 2013; Yap et al., 2017; Tainton-Heap et al., 2021), argues for conserved functions linked to the evolution of central nervous systems, or brains.

In this hypothesis article, we review sleep across phylogeny and propose why some animal brains might need 'active' sleep, in addition to deep or 'quiet' sleep. We examine potential REM sleep functions based on the human and mammalian literature, and then work back from mammals to invertebrates to unpack these functions to some likely evolutionary antecedents. Our hypothesis is that active sleep provides a closed environment for optimizing attention-like processes centered on prediction, ensuring that the real world is predictable enough while maintaining a capacity for surprise. In humans, surprise is associated with emotions, and accordingly REM sleep in humans has been strongly associated with emotional regulation. We propose that this sleep stage has less to do with emotional regulation *per se* and more with an ancient animal need to balance prediction and surprise, in order to be optimally adaptive. We end with a discussion on how active sleep and consciousness might be linked in all animals that have a selective attention and are able to make predictions about what happens next.

PART 1

Rapid Eye Movement Sleep Is Active Sleep

Evidence From Humans

Some of the earliest accounts for sleep and dreaming describe it as either the result of a 'cooling' of the blood during the night or the

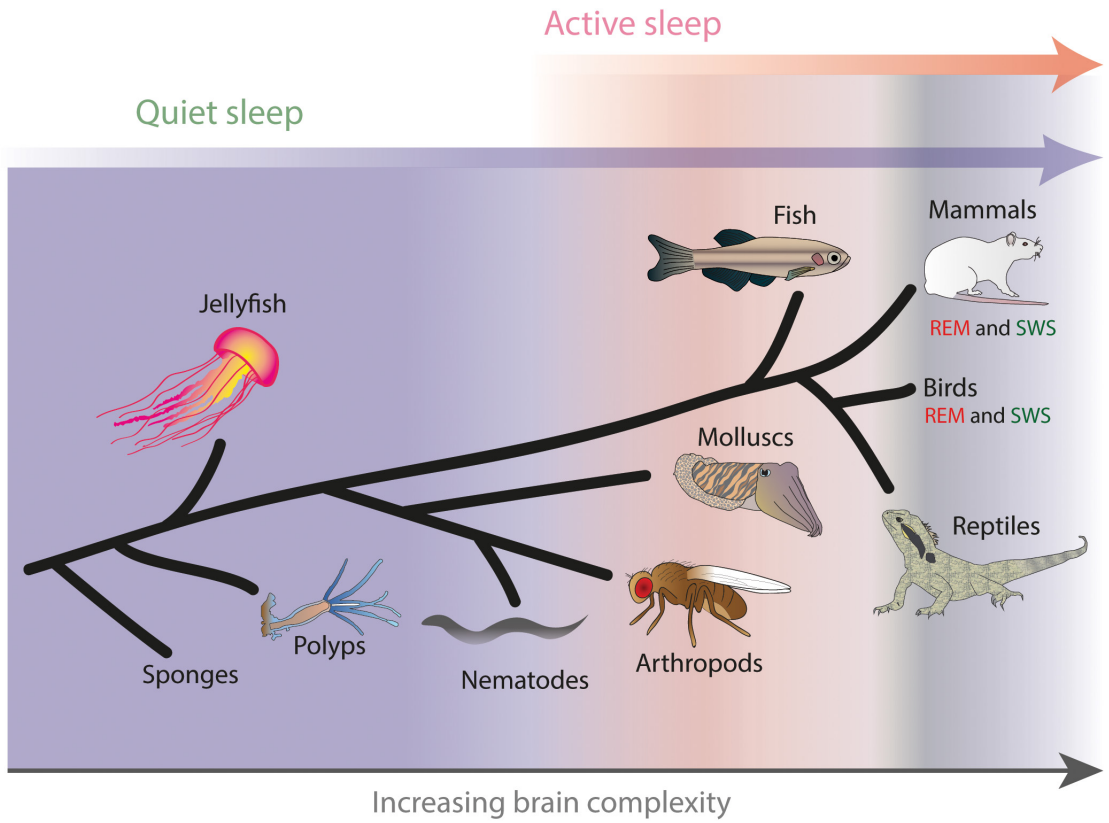


FIGURE 1 | Hypothesized evolution of active and quiet sleep, with rapid eye movement (REM) sleep and slow wave sleep (SWS) in mammals and birds representing specialized solutions to achieving distinct sleep functions. Example animals where different forms of sleep have been characterized are shown, arranged schematically by increasing brain complexity. Adapted from Kirszenblat and van Swinderen (2015).

wicking of an internal fire, while dreams are conjectured to be projections from the divine realm into mortals (Barbera, 2008). Perhaps these theories arose as a way to explain the enforced inactivity and unresponsiveness of sleep, as apart from an obvious continuation of breathing, during this state we appear to others as insensate and immobile. Indeed, this primordial view of sleep as the opposite of activity has led some thinkers to propose that its key function was to keep us safely quiescent in our caves or our trees while predators prowled during the night (Meddis, 1975).

However, over time a number of biological functions have been proposed for sleep, beyond simple inactivity. Before outlining these proposed sleep functions, we first briefly review some important observations about sleep architecture. In humans, a normal sleep cycle consists first of a fairly rapid transition from drowsiness into SWS (usually in the order of minutes). During SWS, the neuronal activity of the brain's cortex is dominated by slow (1–4 Hz) oscillations (termed delta), which have been hypothesized to promote synaptic rescaling (Crunelli and Hughes, 2010; Tononi and Cirelli, 2014; de Vivo et al., 2017; Malafeev et al., 2018). Tellingly, the amplitude of delta activity is greatest at the start of sleep and decreases during successive bouts of SWS throughout the night (Dijk et al., 1990), and the amplitude of these delta waves has been reported to be proportional to the magnitude of sleep pressure experienced by the individual, suggesting homeostatic regulation of processes that accrued during sustained wakefulness (Dijk et al., 1990). That some of these processes involve accrued substances in the brain that need to be normalized after extended wakefulness seemed intuitively obvious; early findings revealed that the cerebrospinal fluid of sleep-deprived animals promotes sleep when injected into waking individuals (Ishimori, 1909; Legendre and Piéron, 1913). This suggested that sleep engages key molecular processes involved in cell health and development. Indeed, more recent studies have associated Non-REM sleep with cell growth and proliferation (Guzmán-marín et al., 2003; Sippel et al., 2020), DNA repair (Cirelli et al., 2004; Vyazovskiy and Harris, 2013; Zada et al., 2019), and waste clearance (Xie et al., 2013). Evidence for these basic cellular functions are supported by observable physical changes in the brain: during the SWS stage of Non-REM sleep, the interstitial space between neurons and glia expands, potentially allowing for more effective clearance of metabolites and other neuronal waste products via the glymphatic system (Xie et al., 2013; Jessen et al., 2015; Fultz et al., 2019). Additionally, glucose usage in the brain is far lower during SWS than during waking, implying that a resetting of local energy stores may be occurring during this sleep stage (Netchiporouk et al., 2001). Thus, the role of SWS in promoting homeostatic cellular processes in the brain is becoming increasingly understood, and

it is intuitively obvious how these processes might have also been best deployed during sustained epochs of inactivity in the first animals.

The predominance of SWS in humans begins to fade typically an hour into sleep, to be replaced by periodic (∼90 min) alternations between SWS and REM (Malafeev et al., 2018). While delta wave amplitude decreases across successive SWS bouts, the duration of the REM bouts increases over sleep time. As with SWS, this may be indicative of homeostatic regulation. However, in contrast to SWS, the functions of REM have proved more difficult to uncover. Nevertheless, in humans REM has been associated with both consolidation of learning (Karni et al., 1994; Boyce et al., 2016) – a function it shares with SWS (Giuditta et al., 1995) – as well emotional regulation (Clemes and Dement, 1967).

REM sleep is also associated with distinct physiological signatures. During REM our constant regulation of internal body temperature (homeothermy) is suspended (Henane et al., 1977; Parmeggiani, 1990). Simultaneously, broad waves of activity originating in the brainstem sweep across the cortex (Jouvet, 1961; Hobson and Friston, 2012), our eyes twitch in their sockets (Aserinsky and Kleitman, 1955; Hong et al., 2009) and penile/clitoral erections are common (Fisher et al., 1965). These are all signs that the brain during REM is active, but in this case, it is internally generated or spontaneous activity rather than responses to the external sensorium. The traveling waves of neural activity (termed PGO waves, for their origin in the pontine-geniculo-occipital nuclei) in particular may resemble evoked visual-like activity in sensory cortices, stimulating wake-like activity (Andrillon et al., 2015; Andrillon and Kouider, 2020). Similarities with waking notwithstanding, REM sleep cannot be simply regarded as a waking state that happens to occur while we are asleep. For example, levels of inhibition in the visual system are lower during REM than wake (Lu et al., 2006; Hobson, 2009b) and the functional organization of the visual cortex and other areas of the brain is more locally confined than during wake (Wehrle et al., 2007). Finally, and perhaps most importantly, arousal thresholds are high during REM sleep (i.e., a strong stimulus is required to return the subject to the 'real' world), and can even be as high as during SWS (Ermis et al., 2010). This raises the question then: Why is a separate stage of sleep needed by the brain, one with wake-like levels of activity? And conversely, what is being accomplished during REM that requires the brain to be largely disconnected from sensory apparatuses?

In science generally an effective assessment of the necessity and functionality of a process is to observe what happens when it is removed or interrupted, although in the question of sleep it can be difficult to disambiguate effects of sleep loss from stress. In the past researchers have experimented with sustained sleep deprivation in humans, finding that perceptual distortions (i.e., hallucinations) were a common result, as well as mood changes and other deleterious cognitive effects (Dement, 1960; Waters et al., 2018). As early as the 1960s it was proposed that the appearance of daytime hallucinations as a result of sleep deprivation was REM-related, as an 'intrusion' of dreams into the waking world (Vogel, 1968). This may be due to the fact that beyond the physiological aspects detailed above, REM is the sleep stage most commonly associated with vivid dreaming (Aserinsky and Kleitman, 1953). While dreams do occur also during Non-REM states, they are typically much less 'dreamlike' and feature a significantly lower number of narrative events (Blagrove, 1993). In contrast, dreams during REM are typically emotionally charged and frequently play upon themes of anxiety or danger (Nielsen et al., 1992). Interestingly, the level of emotional content present within dreams occurring during either early REM or late REM may be predictive of successful emotional regulation (Cartwright et al., 1998). However, for the purposes of this review it should be noted that we are not interested in analyzing dream content, but rather in why this sleep stage should be needed at all. The 'dream pressure' hypothesis (reviewed in Berger and Riemann, 1993) proposes that emotional (and particularly, negative) events generate a 'pressure' to dream that decreases the latency to REM sleep. This may be mechanistically similar to the relationship between sleep pressure and the greater amplitude of early delta waves in SWS, but for emotional content. The implication here is that daytime neurological activity creates a need for cellular homeostatic processes, which are fulfilled by proportional increases in delta activity during SWS, while daytime cognitive or emotional events generate a need for homeostatic regulation by REM sleep. With this hypothesis in mind, we next examine the evidence for active sleep in animals beyond mammals and birds (where they have already been well documented and reviewed (e.g., see Miyazaki et al., 2017; Lesku et al., 2019).

Active Sleep in Other Animals

Active Sleep in Reptiles and Fish

Much like originally in humans, sleep in reptiles and fish has previously been viewed as a simple down-state of decreased brain activity (Siegel, 2008; Libourel and Herrel, 2016), without the delta waves characteristic of mammalian sleep, and without the active paralysis and twitches characteristic of REM. Not surprisingly, this premature conclusion may have been more due to absence of evidence rather than evidence of absence. Importantly, the key neural signatures for identifying sleep stages in mammals are biased toward animals that have a well-developed cortex, the specialized brain tissue capable of generating the kinds of electrical fields that EEG electrodes are designed to detect. This neo-cortical definition of sleep often ignores the rest of the brain, which is largely inaccessible to electrodes placed on the skull's surface. As discussed above, deeper brain recordings into the brainstem of cats and rodents revealed volleys of activity (PGO waves) associated with REM sleep (Jouvet, 1961; Kaufman and Morrison, 1981), suggesting that this more ancient 'reptilian' part of the brain is involved in regulating REM sleep function. It may therefore not have been surprising to discover that reptiles also appear to display a REM-like sleep stage, which alternates with a form of SWS (Libourel and Herrel, 2016; Shein-Idelson et al., 2016). To identify these distinct sleep stages in Australian central bearded dragons (*Pogona vitticeps*), the authors relied on intracranial recordings coupled to filming the reptiles' microbehaviors, such as their eye movements. Instead of specifically identifying neural oscillations such as delta, the authors quantified an ongoing

ratio between high and low frequency domains during sleep and correlated these to the animal's physiology and arousal thresholds. Interestingly, the authors found that bearded dragons appeared to cycle rapidly between sleep stages, with a periodicity of about 80 s (Shein-Idelson et al., 2016). Having identified distinct sleep stages in reptiles, there has so far been little further work in understanding why a lizard might need REM sleep. Examining cognitive functions in lizards is not obvious, as there are few reliable behavioral learning paradigms available.

Fish have been a relative latecomer to sleep research, likely due to the fact that it is difficult to secure electrodes and record electrical activity from unrestrained underwater creatures (but see Ramón et al., 2004; Dunlop and Laming, 2005), coupled with the reliance on brain activity as a readout for sleep. However, with the advent of new techniques researchers in this area have rapidly made up for lost time by exploiting the power of one species in particular, the genetic model *Danio rerio*, or common zebrafish. Following some early observations that freely swimming zebrafish do indeed need to sleep (Zhdanova et al., 2001; Prober et al., 2006; Yokogawa et al., 2007), a breakthrough in assessing neural correlates of sleep in these animals came by exploiting genetically encoded calcium sensors (Chen et al., 2013) expressed in their brain. In a recent study, Leung et al. (2019) imaged the activity of neurons across the brains of sleep-deprived zebrafish that were then restrained for imaging calcium as well as a suite of polysomnography readouts under a microscope. The authors found what appeared to be two distinct forms of brain activity: a putative 'quiet' sleep stage and an 'active' sleep stage (Leung et al., 2019). The former displayed synchronized activity in only a small subset of cells, with most of the rest of the brain becoming quiet. In contrast, active sleep was characterized by volleys of neural activity within the dorsal pallidum, and associated with a number of other physiological readouts (e.g., irregular heartbeat and loss of muscle tone) reminiscent of REM sleep in mammals and birds, but without any rapid eye movements. Together with the earlier behavioral work in this model, these studies support the idea that active sleep has deeper evolutionary roots (and, hence, likely functions) than the mammal-bird-reptile lineage. Importantly, this evidence from zebrafish has spurred the field to move away from neocortical identifiers of sleep stages (e.g., slow-wave sleep and REM sleep) to their likely evolutionary antecedents: quiet sleep and active sleep (**Figure 1**). We therefore next examine evidence for these distinct sleep stages even further down the evolutionary tree, in invertebrates.

Active Sleep in Invertebrates

In our search for distinct sleep stages among invertebrates, it may seem logical to begin with what would superficially appear to be the 'smarter' ones, such as octopi and honeybees. Octopi can plan ahead (Finn et al., 2009), bees can learn context and abstract concepts (Giurfa, 2007), and both use their bodies to communicate complex information with conspecifics (von Frisch, 1967; Young, 1991). Changes in body pigmentation are also evident in relatives of octopuses, such as cuttlefish, and these rapid changes in colors and patterns have also been tentatively associated with emotional states in these animals (Young, 1991; Scheel et al., 2016). Recent work examining sleep in cuttlefish found a behavioral state where the cephalopods were clearly asleep (quiescent and unresponsive) while their pigmentation rapidly flashed a variety of changing patterns, in contrast to other quiescent states where this did not occur (Frank et al., 2012). Without brain recordings, it remained uncertain if this is indeed a form of active sleep, but this has now been confirmed with electrophysiological evidence in a more recent cuttlefish sleep study (Iglesias et al., 2019), as well as in behavioral evidence from octopuses (Medeiros et al., 2021). Importantly, in this last octopus study, careful examination of other microbehaviors allowed the authors to determine transition probabilities between these different sleep states and wakefulness, and these findings further confirm the existence of a complex sleep architecture in invertebrate brains.

Early evidence that sleep architecture might be complex in honeybees relied primarily on filming their microbehaviors in the hive, where they rested. There, it was observed that honeybee antennae moved in a regular, circular pattern soon after sleep onset, and this movement diminished toward the middle of a sleep bout (Sauer et al., 2003), after which the honeybee body lay closer to the substrate, with their antennae drooping and mandibles resting on the surface (Kaiser, 1988). More recent research has confirmed these observations, and shown that bees indeed have deeper and lighter sleep stages linked to changes in microbehaviors (Klein et al., 2014; Zwaka et al., 2015). However, again the absence of electrophysiology (or any other kind of brain recording) makes it difficult to confirm whether these indeed represent 'active' and 'quiet' sleep, as has been documented in vertebrates (but see Kaiser and Steiner-Kaiser, 1983, for evidence of loss of neural responsiveness in sleeping honeybees).

There has been some sleep electrophysiology work done in one unlikely invertebrate, the Louisiana crayfish. In a series of studies performed initially in collaboration with the renowned electrophysiologist Ted Bullock (who recorded from many invertebrates; see Zupanc, 2006), Mexican neuroscientist Fidel Ramón described 'slow' (∼5–10 Hz) oscillatory signatures in the central brain of sleeping crayfish (Ramón et al., 2004). During this sleep stage, crayfish often adopted a stereotypical posture, lying on their side. Subsequent studies from the same group examining these sleep signatures more carefully concluded that local field potential (LFP) activity in sleeping crayfish was not 'slow,' but closer to the beta or low gamma range (15–30 Hz) (Mendoza-Angeles et al., 2010, 2007). Whether this brain activity is always present in sleeping crayfish is unclear, although the authors note that crayfish could adopt other sleeping positions, such as 'crouched' (Mendoza-Angeles et al., 2007). Postural differences during sleep may suggest a form of sleep paralysis, for example the sideways position associated with 15–30 Hz brain activity, but it remains unclear if this is active sleep. As we know from SWS in mammals, neural oscillations do not necessarily indicate wake-like brain activity, which should ideally be verified by neural firing rates. It is nevertheless evident from this work that the arthropod brain does not necessarily become more quiet during sleep, and that sleep-related oscillations seem to emanate from a part of the central arthropod brain termed the 'central complex' (Mendoza-Angeles et al., 2010).

Active and Quiet Sleep in Fruit Flies

The fruit fly, *Drosophila melanogaster*, occupies a special place in sleep research because so much more work has been done on sleep in this model organism over the past two decades, compared to other invertebrates. Sleep was originally identified in *Drosophila* by using re-purposed circadian activity monitors, wherein walking flies interrupting an infrared beam reveal their locomotor activity levels over successive days and nights. Five minutes or more without any beam-crossing was found to be associated with higher arousal thresholds, and thus by inference, sleep (Hendricks et al., 2000; Shaw et al., 2000), and this 5-min criterion was then used for almost all subsequent sleep research in this model, with a view to unraveling the cellular and molecular underpinnings of sleep physiology and function in a simple and genetically tractable model (see Cirelli, 2009; Ly et al., 2018 for recent reviews). This logic held as long as sleep was considered a single state in flies, with a common underlying set of mechanisms and functions. Behavioral experiments probing arousal thresholds more carefully showed that this assumption is unlikely to be true: flies display changing, often cycling, levels of behavioral responsiveness across a sleep bout – deeper sleep and lighter sleep (van Alphen et al., 2013). Further, daytime sleep is significantly lighter than night-time sleep, supporting earlier observations that sleep duration architecture varies between day and night in flies (Ishimoto et al., 2012). More recent behavioral studies using continuous video tracking instead of infrared beams also support the suspicion that flies sleep in different lighter and deeper stages (Wiggin et al., 2020; French et al., 2021; Xu et al., 2021). The realization that sleep might be just as complex in this smallest of animal brains as in higher organisms raises some questions regarding the wealth of correlational data gathered in this sleep model over the past two decades. Indeed, a bewildering variety of neural structures and proteins have been found to be associated with fly sleep (see, for example Dubowy and Sehgal, 2017), if sleep is considered to be a single state based upon a 5-min inactivity criterion. It is now evident that these various structures and proteins probably encompass distinct sleep stages and thus functions, which may have been confounded together. As an analogy, if SWS and REM were confounded in mammals, we would be calling almost every neurotransmitter from acetylcholine to GABA as sleep-relevant and grouping varied structures all the way from the brainstem to the cortex as regulating the same phenomenon, which would obviously be misguided.

Evidence for different sleep stages in *Drosophila* was affirmed with brain recordings in tethered flies walking (or sleeping) on an air-supported ball. The first evidence was electrophysiological, where LFPs recorded from the brains of spontaneously sleeping flies revealed patterns of distinct oscillatory activity alternating with overall decreased LFP activity (Yap et al., 2017). Interestingly, the oscillatory LFP activity was observed to be in the 7–10 Hz frequency range ('theta' band), and was found to emanate from the vicinity of the central complex (Yap et al., 2017; Troup et al., 2018), which aligns with earlier observations from sleeping crayfish – described above. In contrast, 'deep' sleep in flies did not appear to be associated with any specific oscillatory activity, just decreased overall LFP amplitudes (but see Raccuglia et al., 2019 for evidence of 'delta-like' synchronization of neural firing in the central complex of sleep-deprived flies).

Additional support for the idea that flies sleep in distinct active and quiet stages has come from calcium imaging, the same genetic strategy used to identify these distinct stages in sleeping zebrafish. Here, tethered flies placed on an air-supported ball slept spontaneously while a 100 μM volume of neurons in their central brain was imaged using 2-photon microscopy (Tainton-Heap et al., 2021). Tracking the activity of thousands of neurons this way, in waking and sleeping flies, confirmed the complexity previously seen with electrophysiology: brain activity remained wake-like well into the first 5 min of sleep, then decreased to lower levels, and then could increase again to wake-like levels even in flies that remained immobile throughout. Importantly, by tracking the identities of individual neurons throughout a sleep bout, the authors showed that successive active and quiet sleep stages engaged largely non-overlapping groups of cells, suggesting different circuits were recruited and potentially different functions were being served. Indeed, there is now good evidence that active sleep in flies engages a structure in the central brain called the fan-shaped body, which has been linked to sensory processing (Hu et al., 2018; Sareen et al., 2020) and visual learning and attention (Liu et al., 2006; de Bivort and van Swinderen, 2016; Tainton-Heap et al., 2021). In contrast, quiet sleep in flies may be more important for basic cellular homeostatic processes, such as waste clearance (van Alphen et al., 2021) and repair (Stanhope et al., 2020; Bedont et al., 2021). In this way, active and quiet sleep functions in flies may align logically with some proposed REM and SWS functions in mammals, as outlined above.

One conclusion from the admittedly narrow slice of work done in invertebrates suggests that all animals endowed with a brain might sleep in distinct stages, which we propose are best described as active and quiet sleep, and these stages share functional properties with REM and SWS respectively in mammals and birds (**Figure 1**). But what of invertebrates that do not have a brain (or a proper central nervous system), such as sponges, polyps, jellyfish, or certain roundworms? With these, it is possible that only a quiet sleep stage might be operating, tightly linked to periodic developmental or other cell-homeostatic needs (Raizen et al., 2008; Nath et al., 2017; Kanaya et al., 2020). The roundworm *Caenorhabditis elegans* becomes transiently quiescent when growing out of different larval stages (Raizen et al., 2008) or following periods of acute stress (Hill et al., 2014), but there is no evidence (yet) of wake-like levels of neural activity in a quiescent, completely immobile nematode. It is important here to consider recording preparations: calcium imaging in animal models typically requires immobilization of the preparation. While fly or fish brains immobilized under the microscope can co-exist with attached moving legs or tails (to verify sleep or increased arousal thresholds), immobilized nematodes are just that: a worm in plastic straitjacket, unable to move at all. Reports of 'brain' activity during 'sleep' in such preparations (Nichols et al., 2017) should therefore be interpreted with caution, although it remains possible that even these simple animals require periods of active sleep, in addition to quiet sleep.

A case could nevertheless be made for why some animals might not need active sleep, and why all animals might need quiet sleep: not all animals are endowed with a capacity for selective attention (Kirszenblat and van Swinderen, 2015). This debate returns us to our early discussions disambiguating possible REM and SWS sleep functions in humans, with a view to then exploring how some of these functions may have already been required in simpler animals engaged in active sleep.

PART 2

A Role for Rapid Eye Movement Sleep in Emotional Regulation

Having postulated earlier a connection between the wake-like state of REM and emotional regulation in humans, we will now review some evidence for this linkage. We start by discussing the emotion-related effects of altered levels of REM and then move on to ties between REM and common psychological pathologies. In addition, we will briefly review evidence from other mammals where the links between emotional regulation and REM sleep have been modeled and investigated. We then consider how this link might be modeled in invertebrates that display evidence for active sleep.

In insomniacs, REM fragmentation has been linked to emotional dysregulation and an inability to effectively process emotional stimuli (Galbiati et al., 2020) and thus, one attractive option for investigating the functions of REM is to observe the effect of its removal in normal and pathological subjects. With the advent of polysomnography and online analysis of EEG data, it has become increasingly tractable to accurately identify waking and sleeping states of experimental participants and to use this information to selectively interrupt specific sleep stages. For REM in particular, numerous studies have shown that it is involved in recall of emotional content (Nishida et al., 2009; Rosales-Lagarde et al., 2012; Wiesner et al., 2015). In the work of Rosales-Lagarde et al. (2012) it was shown that human participants deprived of REM were less able to accurately distinguish between trained and novel images containing negative emotional content but were unimpaired in their recall of emotionally neutral stimuli (Rosales-Lagarde et al., 2012). Wiesner et al. (2015) also utilized selective deprivation of both SWS and REM, showing that emotional memory consolidation (quantified as successful recognition of stimuli on the following day) was impaired by REM deprivation but emotional reactivity (self-reported on a survey) was unchanged between the deprivation groups (Wiesner et al., 2015). The implication here is that SWS may contribute to the regulation of emotional reactivity, while emotional consolidation is primarily controlled by REM.

A role for REM sleep in emotional memory consolidation can also be found through fear conditioning studies, as an alternative means to access emotion. Spoormaker et al. (2012) conditioned human subjects (with mild electric shocks) to feel fear toward simple visual shapes. These subjects then underwent extinction training (presentation of the shapes in an absence of the aversive shocks) after fear conditioning and were split into either a REM deprivation group or a group with a matched amount of Non-REM deprivation. It was found that REM deprivation significantly impaired the effectiveness of the extinction training, with REM deprivation participants exhibiting responses to conditioned stimuli closer to original than post-extinction levels (Spoormaker et al., 2012). This shows that REM sleep plays a key role not only in forming associations between emotional events and their eliciting stimuli but also in the weakening of such ties when applicable. In humans and other mammals, processing of emotional events during REM is proposed to revolve around activity in the amygdala and anterior cingulate cortex, such that impairment in normal functioning prevents effective emotional consolidation (Braun et al., 1997).

In rodent models of fear conditioning, it has been possible to begin investigating more carefully the links between REM and emotional learning. One reliable technique for selective REM deprivation in rodents involves a semi-submerged sleeping platform where Non-REM sleep (which does not require muscle relaxation) can be achieved but REM onset leads to sudden immersion and awakening (Arthaud et al., 2015). Early behavioral work using this approach showed that rats deprived of REM sleep had decreased acetylcholine levels (Bowers et al., 1966) and were more prone to fight following an unexpected foot shock (Morden et al., 1968). Conversely, it has been shown in mice that fear conditioning can lead to an increase in REM sleep in subsequent rest periods (Smith, 1985). In more recent rodent work it has been shown that muscarinic cholinergic receptors are critical for REM sleep (Niwa et al., 2018) and knockdown of cholinergic receptors significantly impairs fear conditioning, as well as other forms of learning (Queiroz et al., 2013). However, acetylcholine regulates a wide range of waking brain functions, so it is difficult to draw any strong conclusions between learning and REM sleep without considering other consequences of chronically downregulating cholinergic receptors in the mammalian brain.

One aspect through which ties between REM and emotional consolidation become salient is that of pathological symptomologies. In particular, the association between REM and depression is arguably the most classic neuropathology of negative affect (Berger and Riemann, 1993). Sleep studies with clinically depressed subjects have been performed since the 1940s (Diaz-Guerrero et al., 1946) and in these studies and the decades since it has been shown that depressed individuals tend to have reduced volume of SWS and shortened latency to enter REM sleep (Berger and Riemann, 1993). However, Vogel et al. (1980) showed that the total volume of REM was not significantly different between depressives and neurotypical individuals, and the change in REM architecture was primarily a shift toward 'early REM' in afflicted individuals. In more recent work, Harrington et al. (2018) recruited participants with minor and severe depression, finding that the degree to which participants consolidated new negative memories during a night of sleep was correlated with the severity of their depression and an increase in REM density. Notably, while there is evidence that REM deprivation leads to emotional instability (Clemes and Dement, 1967), there is also significant evidence supporting the use of selective deprivation of REM as a tool leading to improved outcomes for sufferers of depression (Vogel et al., 1980),

although more recent evidence has cast doubt upon the REM specificity of this improvement (Giedke and Schwärzler, 2002). Many commonly prescribed antidepressants [such as the older tricyclic and tetracyclic antidepressants as well as more modern selective serotonin reuptake inhibitors (SSRIs)] have a REM-suppressing effect (Reyes et al., 1983; Riemann et al., 1990). One explanation for these seemingly contradictory findings is that both too much or too little REM is deleterious to normal emotional functioning, so these REM-suppressing antidepressants may improve depression by returning the latency of REM onset to a normal point (Reyes et al., 1983). There have also been other propositions for the mechanism behind emotional improvements following REM deprivation, ranging from a resetting of a biological oscillator (Vogel et al., 1980), to prevention of dreams containing negative emotional content during early REM epochs (Cartwright et al., 1998).

However, changes in REM quantity are not just associated with depression; other neurological disorders including post-traumatic stress disorder (PTSD) (Yetkin et al., 2010), obsessive-compulsive disorders (OCD) or eating disorders (Berger and Riemann, 1993) and schizophrenia (Zarcone et al., 1987) have all been linked to alterations in this sleep stage. But why might REM be increased in patients with these diseases in the first place? One possibility is that depression, PTSD, schizophrenia, OCD and other cognitive disorders are different manifestations of similar underlying neuropsychological issues (Plana-Ripoll et al., 2019; Hobson et al., 2021), or alternatively that the dysregulation of emotional content invariably involves a REM element. This "chicken or egg?" question is centralized around whether it is the disorders that lead to dysregulated REM sleep, the dysregulated sleep that leads to disorders, or a combination of both possibilities. The difficulty of determining whether altered REM architecture is a cause or consequence of cognitive and emotional disorders calls for a reductionist approach where key aspects of REM sleep, such as emotional regulation, might be modeled. Although there is no evidence that REM sleep evolved from invertebrate active sleep, the discovery of active sleep in a variety of simpler animal models provides a way forward for understanding potentially conserved sleep functions. However, with the evidence from humans and rodents pointing so strongly toward emotional regulation, how can this even be modeled in animals such as flies?

Emotions in Arthropods?

There have been a few published efforts to determine whether arthropods display emotional responses. Although emotions seem to be largely subjective, thus opaque to anyone beyond ourselves, they also betray a short list of clearly measurable correlates which can be used as potential evidence. These correlates are centered around measures of arousal or bodily excitation, as well as evidence of valence, which can lead to attraction or repulsion to a stimulus. To uncover any evidence of a persisting 'internal' state, behavioral responses are then often dissociated from immediate stimulus parameters. For example, positive or negative valence might generalize to related stimuli or graded variations of the stimulus, or altered arousal states might persist well after the stimulus has disappeared (Anderson, 2016). Such criteria have been useful for studying aggression in a wide variety of arthropods, from crayfish to flies (Kravitz and Huber, 2003). Lean explanations of emotions however might view aggression as an innate response, much like phototaxis or courtship. To probe more deeply into learned emotional responses (e.g., something innate might be overturned after learning new associations), researchers have traditionally resorted to classical conditioning paradigms, by punishing or rewarding animals and then designing elegant experiments to see if some of the emotion-relevant criteria (e.g., scalability, persistence, generalization) are satisfied (Anderson, 2016). Typically, these experiments are designed to determine if animals are behaving 'optimistically' or 'pessimistically' when confronted with ambiguous stimuli, after positive or negative re-enforcement. For example, crayfish (*Procambarus clarki*, the same species discussed earlier) was found to display anxiety-like behavior after punishment (Fossat et al., 2014). Remarkably, this behavior could be corrected with the anti-anxiety medication chlordiazepoxide, developed originally for humans (Fischer et al., 2006). Similar experiments on honeybees showed the same result, with these clever insects displaying a form of pessimism about ambiguously colored flowers after being shaken (Bateson et al., 2011). Conversely (in a different study), when bumblebees received an unexpected reward immediately prior to performing a feeding choice task they were more likely to display 'optimistic' behavior by promptly moving toward ambiguous stimuli that control bees were slower to attend to Solvi et al. (2016). To gain traction, these behavioral studies often support their conclusions with pharmacological interventions, typically centered on drugs targeting monoaminergic systems such as dopamine and serotonin, which also regulate emotional responses in mammals (Anderson, 2016).

Any neuroethologist attempting to uncover evidence for emotions in insects is, however, confronted with a conundrum: we could in principle document bumblebees sobbing in grief at the death of a conspecific, tiny handkerchiefs and all, and a counterargument could always be made that this is nothing more than a series of innate behaviors, not emotion. This potential stalemate has led some in the field to take a different tack, that is to use reductionistic models such as *Drosophila* to simply better understand the neural underpinnings of arousal and the brain circuits regulating the variety of behaviors that might provide mechanistic evidence for scalability, persistence, and generalization. Thus, one *Drosophila* study (Gibson et al., 2015) designed a visual threat paradigm to measure defensive arousal in flies ('fear'), hinting at the existence of dynamic internal states. Other studies have uncovered evidence of efference copy mechanisms (Blakemore et al., 2000) in the fly brain, suggesting that internal states (or motivations) gate the responsiveness of sensory neurons (Kim et al., 2015; Fujiwara et al., 2017). A recent study provides some additional convincing evidence for internal states in flies, by probing how visual processing might be dynamically gated by sexual arousal (Sten et al., 2021). If sexual arousal gates visual processing in flies, it seems likely that fear or anxiety might too, although there has not been much work done in unraveling the neural circuitry of fear in flies. In contrast, there has been much circuit-level work done

on aggression (Hoopfer, 2016) and escape responses (Card and Dickinson, 2008; Fotowat et al., 2009), without any need to invoke emotions like anger and fear. This brings us back again to our original conundrum of how to disambiguate emotions from innate responses in these simpler models, and more specifically how to disentangle our anthropocentric views of emotion from their likely evolutionary antecedents. One way to proceed in this regard, and also to potentially better understand conserved functions being engaged by active sleep, is to study selective attention mechanisms and to consider how emotions are linked to predictive systems in the brain.

Like humans and rodents, insects pay attention to novelty. This means that, when confronted with novel objects [in a virtual reality environment, for example (Heisenberg and Wolf, 1984)], flies will orient toward objects they haven't seen before and ignore competing objects they may be more familiar with (Dill and Heisenberg, 1995; Solanki et al., 2015). Interestingly, responses to visual novelty in flies can override innate visual preferences, meaning that flies will transiently fixate on innately 'repulsive' objects (e.g., a green square) over 'attractive' objects (e.g., a vertical green bar) if the otherwise repulsive object is novel (Grabowska et al., 2018). Earlier electrophysiological recordings from behaving flies showed that visual novelty is associated with transient oscillations in their central brain, in the range of 20–30 Hz (van Swinderen and Greenspan, 2003; van Swinderen, 2007). A more recent study recording directly from the central complex of behaving flies revealed a selective phase-locking mechanism between the endogenous 20–30 Hz oscillations and the attended object (Grabowska et al., 2020). This suggests a conserved mechanism in the fly brain attuned to first detecting surprising stimuli (i.e., novelty), and then to paying attention to them for a period of time (Sareen et al., 2011; van Swinderen, 2011). Interestingly, when an arousal system in the fly (neuropeptide F) is transiently activated, this increases 20–30 Hz phase locking in the fly brain and redirects the insect's attention to novel objects irrespective of their innate valence (Grabowska et al., 2020). Such findings again suggest an evolutionarily conserved link between arousal systems and novelty detection mechanisms. To further consider this link with predictive mechanisms in the brain, and how they might be regulated by active sleep, we return to humans.

PART 3

Emotions Are Linked to Prediction Errors

There is an obvious purpose to emotions, which is to alert us about the consequences of our predictions. Unfulfilled predictions are jarring; we might feel sadness or anger when our favorite sporting team unexpectedly loses a match, or more acutely when we miss a goal kick. Similarly, there is a simpler satisfaction when a prediction is confirmed (for better or worse). In this way, emotions are a way to inform us that a salient event that failed to match our predictions has occurred, and that the circumstances that lead to this should be corrected and committed to memory. Numerous psychological studies have shown a relationship between the strength of emotional responses and the degree to which events were surprising (e.g., Feather, 1967; Bhatia et al., 2019). Notably, emotions seem to arise more from the deviation of expectation of an event rather than the magnitude of the event itself. In work by Villano et al. (2020), it was shown that for university students receiving their end of semester course marks, the strength of emotional affect experienced by the students was more strongly proportional to the deviation from their expectations than the mark itself. Additionally, and perhaps unsurprisingly, negative affect (resulting from lower than expected marks) was more profound than positive (Villano et al., 2020). These examples of high-level cognitive predictions are what we typically think of when associating emotions to surprising events. However, predictions can also reflect low-level (non-explicit) expectations, and these can also trigger emotional responses that might be rationalized afterward.

Recent theories seek to explain emotions as a way to understand both explained and unexplained deviations in our own internal state (Seth, 2013; Barrett et al., 2016). For example, Schachter and Singer showed in 1963 that participants who were administered an injection to increase their physiological arousal (in this case, a low dose of epinephrine), but not informed as to the effects of said injection, were more prone to sympathetic emotional influence from a conspirator who had been schooled to act in a particular emotional manner (Schachter and Singer, 1963). The implication here is that in the absence of their internal narrative providing them an obvious cause for their self-detected state of arousal, participants attributed their internal state as the result of a presumed emotional reaction. Similar evidence can be found in the classic psychological quirk of mood improvement following a pen being held in one's mouth to artificially induce a smile (Labroo et al., 2014). Experiments such as these could be seen as attempts at divorcing emotional responses from the conscious states typically associated with them in humans, to potentially achieve a better understanding of their fundamental functions. One of these potential functions is to highlight that something unexpected has just occurred, which introduces predictive coding theory to our discussion.

Predictive coding theory (Rao and Ballard, 1999) provides a compelling framework on which to better understand the importance of emotional regulation, based on notions of 'unconscious inference' first proposed by Hermann von Helmholtz (Helmholtz, 1860; Shipp, 2016). Predictive coding describes a system whereby sensory information about the world is used to generate an internal model that informs a system about the likely causes of said sensory stimuli. Sensory returns not matching this model represent prediction errors and the system can react to these by updating its model to better fit the evidence or by enacting change to bring the world into line with the model (**Figures 2A,B**). For these models to remain efficient and parsimonious, it is necessary for them to be regularly reviewed and reorganized, which is a role some have proposed for REM sleep (Hobson et al., 2014; Llewellyn, 2016; Windt, 2018).

In humans, predictive processing is commonly studied through the optics of 'oddball' paradigms, wherein a sequence of 'standard' stimuli is interrupted infrequently by a 'deviant' stimulus (Friston, 2005; **Figure 2C**). Under normal conditions

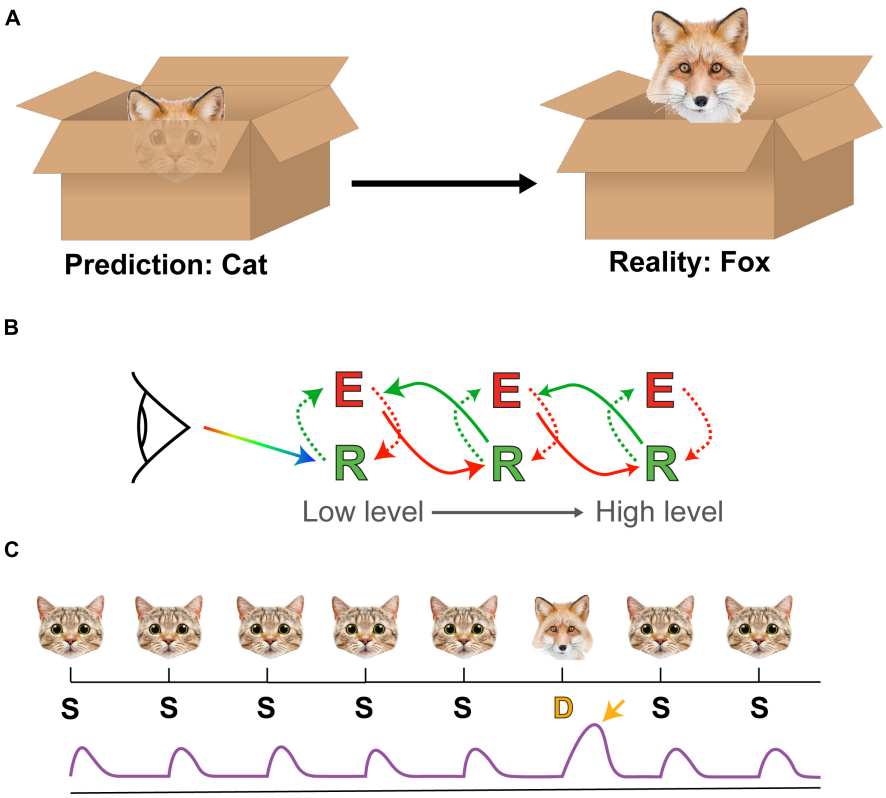


FIGURE 2 | Predictive coding and oddball paradigms. **(A)** A prediction of the animal hiding in a box (a cat) based on a set of ears turns out to be in error (it's a fox) when further details about the animal are revealed. In this case the prediction error is the misattribution of the animal as a cat. **(B)** A simple schema of core tenets of predictive coding theory. Sensory input (rainbow arrow) interfaces with a low-level representation (R) unit, which generates a mismatch that is used to refine an error (E) signal within a feedback architecture. This error signal also receives predictions from higher-level representation units while simultaneously supplying these units with updates. By arranging these units in a hierarchical manner, each layer can be used to represent different levels in processing, all the way from simple visual features such as orientation up to abstract concepts and ideas. **(C)** A schema of a simple oddball paradigm and prediction error signal. In this case an image of a cat (the Standard, S) is presented repeatedly, occasionally replaced with an image of a fox (the Deviant, D). The standards (S) evoke a reproducible response from the brain (purple trace) while the deviant (D) (typically matched for low-level features) evokes a different response (yellow arrow), which is detectable by EEG and/or functional magnetic resonance imaging (fMRI).

this deviant stimulus elicits a prediction error signal in EEG recordings, visible in humans as a Mismatch Negativity (MMN), which is a distinct electrophysiological correlate of surprise (Näätänen, 1990). The usefulness of oddball paradigms lie in their versatility; virtually any sensory modality can be used for delivering stimuli and the semantic separation between standard and deviant stimuli can be as simple as "square vs. circle" (Huettel and McCarthy, 2004) or as complex as "repeated human face vs. novel human face" (Feuerriegel et al., 2018).

The Sleeping Brain Makes Predictions

Notably, the human brain appears capable of generating certain prediction error signals even during the various stages of sleep. Previous studies have shown that human participants elicit electrophysiological markers of surprise in response to deviant stimuli during waking, Non-REM and REM sleep (Bastuji et al., 2002) and at least one study has shown that the rate of K-complexes during sleep may be tied to the salience of presented stimuli (Oswald et al., 1960). Notably, high-level signals of prediction error such as the P300 wave [so named because it is elicited around 300 ms after recognition of a deviation (Picton, 1992)] do not occur in response to oddball events during either Non-REM or REM sleep, but local detections of mismatch are present (Strauss et al., 2015). These prediction error signals have also been studied in the context of altered brain states such as general anesthesia (Koelsch et al., 2006) and coma (Bekinschtein et al., 2009), wherein the local mismatch response is typically preserved whereas more 'conscious' indicators of deviation fail to arise. Recognition of one's own name, which has been long known to occupy a privileged space in human stimulus processing (Carmody and Lewis, 2006) is present even in sleep (McDonald et al., 1975), implying that it is a representation that may span all the way to the lowest levels of the auditory system. Thus, while the sleeping brain still appears to be able to categorize external events as surprising or not surprising, it remains unclear to what level different sleep stages regulate this important capacity of the brain.

Studies have shown that sleep in general seems important to the formation of predictive models (Wagner et al., 2004; Lutz et al., 2018). For example, improvements in prediction-associated performance were found by Wagner et al. (2004) on a digit

transformation task with a hidden abstract rule. Under normal conditions participants would derive an answer for each task block by stepwise calculations, but it was also possible to infer the correct answer midway through each block if participants were to discover the hidden rule governing the digits, the existence of which was not communicated to participants. Comparing participants who were allowed an 8-h sleep against those who remained awake revealed that sleepers had a more than doubled likelihood of uncovering the hidden rule the following day, compared to participants who were awake for the same span (Wagner et al., 2004).

Does Rapid Eye Movement Sleep Specifically Regulate Predictions?

Understanding that emotions provide a potential mechanism to recognize and correct prediction errors, and that REM sleep is involved in emotional regulation, immediately suggests that REM sleep might also be important for regulating prediction. Thus, we posit that rather than regulating emotions *per se*, REM sleep in fact regulates the *predictions* that drive our (human) emotional responses. Importantly, this view allows us to sidestep anthropocentric concerns on whether animals have emotions or not; they all make predictions.

The evidence for REM involvement in consolidation of learned tasks is extensive, but arguably the end goal of a consolidated model is for it to be actively used in a predictive capacity and, so far, human experimental literature directly linking together REM sleep specifically with predictive capacity remains relatively unexplored. Barsky et al., 2015 tested participants unconsciously learning to predict the 'weather' from hidden association probabilities with abstract stimuli before and after a nap. They found that the nap significantly improved participant's ability to correctly guess the weather, and that REM quantity was correlated with success (Barsky et al., 2015). Earlier work by Cai et al. (2009) showed that REM sleep specifically was important for improvement in creative problem solving, involving recombination of learned sequences with an unrelated cognitive task. Notably, the brain is capable of forming models of stimulus properties even without conscious direction (Barbosa and Kouider, 2018), which is probably one component underlying the means by which a sleep state like REM can have such a seemingly cognitive role. For paradigms targeting unconscious aspects of prediction, responses to certain unpredictable 'oddball' stimuli can be found during REM (Atienza et al., 2000; Sculthorpe et al., 2009), but not Non-REM (Cote et al., 2001; Sabri and Campbell, 2005; see Ibáñez et al., 2009 for a review), implying that during REM sleep the brain is in a state conducive to the evaluation of predictive models.

Some stronger evidence for a connection between predictions and REM sleep comes from ties between REM and activity in the hippocampus, a structure in mammalian brain associated with working memory (Tesche and Karhu, 2000). So-called 'place cells' in the hippocampus have been shown to encode specific and unique regions of physical space (O'Keefe and Dostrovsky, 1971), making them a prime candidate for processes involving consolidation of predictive models. Hippocampal replay of place cell firing sequences has been shown in rats (Lee and Wilson, 2002) and other animals (Ulanovsky and Moss, 2007) during both SWS and REM sleep. Interestingly, hippocampal replay is commonly associated with theta-band (4–8 Hz) activity during REM (Tesche and Karhu, 2000). As an endogenous rhythm, theta seems critical to the process of memory consolidation within the mammalian hippocampus (Cote et al., 2001), and 'phase precession' mechanisms (**Figure 3**) appear to be a key feature linking diverse firing sequences into a compact predictive code defined within different theta oscillation cycles (Jaramillo and Kempter, 2017). In essence, rather than being just a simple rate code, wherein different cells fire more when animals cross certain physical spaces (**Figure 3A**), each successive space is actually anticipated (due to past experience) as a unique firing sequence within a theta cycle (Jaramillo and Kempter, 2017). In this way, the hippocampus is able to encode information into the theta band activity at a timescale that is also conducive to spike timing dependent plasticity (STDP), meaning that confirmed predictions are strengthened and thus preserved as firing sequences (D'Albis et al., 2015), whereas prediction errors might jolt the system into a new coding sequence. It seems probable that a whole range of neuronal firing events are temporally organized within successive theta cycles, creating an opportunity for strengthening links among a variety of modalities and memories, not just sequential physical spaces. By reactivating theta (and thus, the predictive information provided by the aforementioned phase precession sequences) during REM sleep (Lee and Wilson, 2002; **Figure 3B**), the brain is able to effectively revisit these temporal sequences and regulate their synaptic strengths (Skaggs et al., 1996). It seems intuitive to extrapolate from this observation that such a role for REM sleep in optimizing predictions about physical navigation through space might generalize to other predictive capacities, such as sensorimotor or social.

In humans, Karni et al. (1994) showed that disruption of REM sleep impaired performance on tasks learned immediately prior to the REM deprivation but not on previously learned tasks, and when Non-REM sleep was disrupted there was no impairment to performance. Similar studies performed in mice have shown that the theta rhythm present during REM sleep is a critical component of this new-task consolidation (Boyce et al., 2016), probably through reactivation of neurons phase locked to the theta cycle. Given that theta is absent during SWS but present in wake and REM (Green and Arduini, 1954), it seems logical to infer that one aspect of REM may be engagement of wake-like processes to reorganize place cell activity and thereby allow the brain to build better predictive models. It is unknown, however, if invertebrates display predictive processes such as phase precession, but it is interesting to note that active sleep in *Drosophila* flies (Tainton-Heap et al., 2021) seems to be characterized by a theta-like (7–10 Hz) oscillation (Yap et al., 2017).

One potential clue that active sleep is associated with building predictive models comes from sleep ontogeny, or how sleep architecture changes through life. Most young animals need more sleep than adult animals (Kayser and Biron, 2016). In contrast, sleep is significantly reduced in old age, although this can

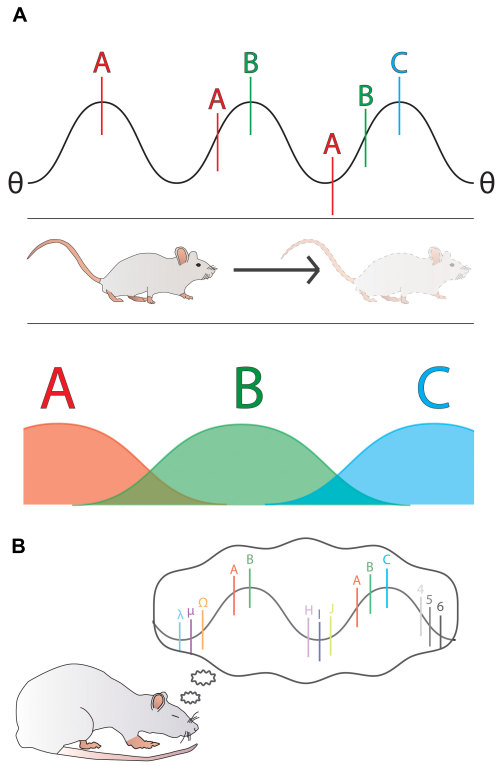


FIGURE 3 | A general overview of phase precession as represented by an example of a rat moving through 1-dimensional space. **(A)** A hippocampal theta (θ) waveform has overlaid onto it the activity of three place cells (spiking frequencies are represented as A, B, and C; bottom). The position of each neuron's spiking on the phase of theta is determined by the temporal sequence of the rat's movement. As the rat moves through the regions represented by A, B, and C, the phase of each of these neuron's spiking shifts further from 360°, such that each successive space is represented by an ordered sequence within a single theta cycle. **(B)** By reactivating theta during REM sleep, a rat replays the temporal sequences that became phase locked to theta during waking. It can be seen that the activity of many more neurons than A, B, and C could be encoded and linked onto theta, representing the role of theta in encoding more than just place fields and thus creating a variety of predictive models.

be harder to disambiguate with encroaching neurodegenerative conditions such as Alzheimer's and Parkinson's, which are co-morbid with impaired sleep (Okawa et al., 1991). When partitioned between REM and Non-REM, it becomes clear that most of the change in sleep architecture through life (at least in humans) can be attributed to decreased REM, with this active sleep stage accounting for almost a third of a newborn's life and only ∼5% of an elderly individual's time, while Non-REM sleep duration stays comparatively more constant (Roffwarg et al., 1966). Intriguingly, REM sleep has been shown to occupy an even greater proportion of prenatal life, when infants are still developing in the womb (Peirano et al., 2003), with some proposing that early human brain development may be almost entirely REM-like (Coons and Guilleminault, 1982; Hobson and Friston, 2012). The observation that prenates and infants display substantially more REM sleep could suggest that this sleep stage has less to do with dreams *per se* (what might prenates dream about anyway?) and more to do with satisfying key needs of developing brains, such as neural reorganization (Cao et al., 2020). Following from our discussion above, one important need appears to be optimizing the capacity to make predictions about one's actions, and thereby build models about one's own body plan. Notably, human studies of proprioceptive efference copies have indicated a modulating role for theta oscillations (Stock et al., 2013), a role which would align well with the preponderance of REM in early brain development. It would seem reasonable to propose that most learning in the womb is proprioceptive, namely concerned with establishing control over different body parts and determining what sensory events have internal versus external causes. As newborns develop, other predictive models more relevant to life outside the womb need to be developed, and this ongoing need to learn, with perhaps a matching need for REM, continues through childhood but wanes in adulthood. Although not an explanation for REM sleep, this correlation with sleep ontogeny provides a powerful entry point into potentially exploring active sleep in non-human animals, from other mammals to invertebrates. This is because such an explanation sidesteps any need to explain dreaming (what does it matter what prenates – or flies – might be dreaming about?) and focusses instead on functional explanations linked to optimizing predictive models – something highly relevant to most motile creatures that have to anticipate the consequences of their actions.

DISCUSSION

The evolution of sleep and attention is probably intertwined (Kirszenblat and van Swinderen, 2015), and here we propose that it is active sleep specifically that has co-evolved with animals' capacity to pay attention to surprising events in their environment. Whereas quiet sleep (or SWS in mammals and birds) is increasingly found to be associated with homeostatic repair processes that collectively appear to be attempts at reducing cellular entropy in the brain following waking activity, active sleep may instead reflect cognitive homeostatic process aimed at optimizing how animals predict the world. This hypothesis has interesting implications for the evolution of subjective awareness across animals, and for the role of active sleep in curating this capacity throughout the life of individual animals. Specifically, we propose that the ongoing debate on the origins of consciousness in animals (Barron and Klein, 2016) could be productively informed by understanding which animals have evolved a need for active sleep alongside quiet sleep.

Brains could be viewed as evolving prediction machines. We discussed earlier how emotional responses associated with prediction errors might be important for forming new memories, to enable brains to become even better prediction machines. Thus, a joke is typically funny the first time because of some unexpected twist, but rarely funny the second time: we predict the twist. Other than humans, animals don't seem to joke much, but most animals are probably well tuned to detect prediction errors more relevant to their individual niches. Most animals might make use of endogenous arousal and valence systems to

detect prediction errors, and thereby highlight the need for an updated prediction. Yet, this process needs to be finely tuned. Too many prediction errors might indicate a maladaptive inability to generalize, while too few prediction errors might result in an inability to learn anything new. Herein lies a paradox: the mechanism that brains seem to employ to detect and correct prediction errors (emotion, or arousal) is the same quality that brains are attempting to eliminate by becoming better prediction machines.

Indeed, this paradox has been discussed in machine learning and philosophy. For example, regarding the difference between novelty and surprise in computational neuroscience (Barto et al., 2013; Schwartenbeck et al., 2013), or in the 'dark room' problem in philosophy (Friston et al., 2012), which puts forward the following conundrum: if brains are designed to minimize surprise, then why don't animals act to minimize unpredictable events by seeking environments that remove certain stimuli entirely (e.g., a dark room)? A resolution to this paradox has been proposed at the level of predictive coding theory: the minimization of prediction errors in the moment could be viewed as fundamentally different from choosing actions that will minimize prediction errors in the future. Technically, prediction errors correspond to surprise, while 'expected' prediction errors – consequent on action – correspond to uncertainty. This follows because surprise is self-information in information theory and expected surprise is entropy or uncertainty. Thus, there is a key difference between a surprising event that was unpredicted and choosing an action that you expect to bring about unpredictable outcomes. Minimizing expected surprise is the tenet of active inference and rests upon a good generative or internal model of the consequences of actions. A role for sleep in this setting has been proposed previously (e.g., see Friston et al., 2017). Active sleep could provide an opportunity for the brain to simulate and test a broad range of internal models, which is probably a more adaptive strategy than seeking a metaphorical dark room of zero surprises.

Imagine a brain becoming so good at predicting everything in its environment that it never becomes surprised anymore, and thus never evokes an emotional response to highlight a prediction error. Such a brain might not be very different from a computer: just an input/output system working within an invariant universe. Such a brain would not need emotion, since in a world of perfect predictability there is no surprise and thus no need to consolidate new memories. Indeed, it might be doubtful whether such a brain would be conscious, in the way that term pertains to subjective experience (Barron and Klein, 2016). A brain moving toward zero surprises might sound adaptive, but it probably isn't. This is because the world is never entirely predictable. A brain in a closed environment (e.g., a baby in the womb, a monk in a monastery, or a fly in a bottle) may achieve close to perfect predictability in that specific environment, but this does not do it any good outside that environment. We are always surprised, because our world is always changing, and this requires continuously updating our models of the world. This is important from the point of view of cognitive flexibility and adaptability.

Cognitive flexibility comes hand-in-hand with minimizing redundancy and maintaining a degree of latitude when forming accurate accounts of the (waking) sensorium. One view of active sleep that speaks to this imperative builds on ideas from statistics and machine learning (Hinton et al., 1995). In this setting, the maximization of model evidence entails a minimization of statistical complexity. This can be seen from many perspectives. For example, in the free energy principle proposed by Friston et al. (2006), the implicit maximization of entropy is one way of ensuring that we keep our options open when forming beliefs about states of affairs in the world (Hobson and Friston, 2012). This may seem in opposition to proposed deep sleep functions, which are aimed at decreasing entropy or complexity in the brain, which has been formulated in the context of minimizing synaptic connections (Tononi and Cirelli, 2006). It is possible that active sleep – and the rehearsal of narratives and contingencies accumulated during the day – is similarly in the service of removing redundant connections and thereby minimizing complexity. Cognitive flexibility could thus be seen as emanating from processes that preclude overfitting overly parametrized internal models (with redundant and exuberant synaptic connections) (Hoel, 2021). This view would tie neatly with the synaptic homeostasis hypothesis that has been attributed to deep sleep in higher animals (Huber et al., 2004; Tononi and Cirelli, 2006).

A related view however might be that the neural reorganization that seems inherent to active sleep (Cao et al., 2020) ensures that the crucial cellular repair/homeostatic processes engaged during deep (quiet) sleep do not compromise cognitive flexibility. Thus, what begins as a necessary model-building exercise during brain development persists throughout life (albeit often to a lesser degree; Herman et al., 1991; Hobson, 2009a), as a crucial mechanism for maintaining cognitive flexibility. By drawing links among events (or neuronal groups) which would not ordinarily be associated in waking life, active sleep might ensure that valence systems (how value is assigned) remain tuned at an optimal level, allowing for an appropriate level of surprise while awake. One way to do this may be to disconnect the waking brain from the outside world for extended periods of time. In this sense, a key function of active sleep – in any animal – may be to entertain a quasi-infinite range of alternate possibilities (by replaying or remixing neural sequences, as in **Figure 3B**), to ensure the waking brain remains just enough surprised about the real world to keep paying attention and learning. Consciousness is thus adaptive, but it doesn't come for free. We need to dream to keep from becoming habit-driven, entropy-minimizing robots.

While the link between attention and consciousness remains debated (e.g., see De Brigard and Prinz, 2010; van Boxtel et al., 2010), a focus on optimizing prediction provides an effective strategy to investigate a role for active sleep in simpler animal models such as flies. In predictive processing and active inference, attention is usually described as assigning greater precision to certain sensory streams or posterior beliefs (Feldman and Friston, 2010). Simply put, precision in this instance is an estimate of predictability. Physiologically, it is thought to be encoded by neuromodulatory mechanisms that control synaptic gain (Kanai et al., 2015). Thus, assigning precision in a context-sensitive fashion (i.e., cognitive flexibility) looks very much like attention.

The key point here is that exactly the same neuromodulatory mechanisms that underwrite attention – and the deployment of precision during hierarchical predictive processing – are those thought to be responsible for active sleep and dreaming (Hobson, 2009b). This speaks to our notion that dreaming and attention may inherit from the same (classical) neuromodulatory systems.

The idea that dreaming might shape our consciousness is not new (Hobson, 2009a; Hobson et al., 2021; Windt, 2021). What is new is the realization that many other animals, including even flies, seem to have an active sleep stage. This suggests that something more primordial than consciousness is being attended to by periodically uncoupling a waking brain from the outside world. This view implies that this primordial quality is adaptive, meaning that it helps animals survive. This view also suggests that this trait might be a feature of all animals that show any evidence of active sleep. We propose that what is being curated here is a balance between prediction and surprise, which shapes how animals pay attention. Rather than being a simple indicator for which animals are conscious and which are not, we propose this as an effective strategy to understand how subjective awareness may have evolved from such a mechanism. It will for example be interesting to verify the extent of active sleep across the animal kingdom and see how this might correlate with different animals' capacity to optimize prediction error signals.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

MVDP and BVS wrote the manuscript. Both authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Australian Research Council DP210102595 to BVS and a University of Queensland Research Training Scholarship to MVDP.

ACKNOWLEDGMENTS

The authors would like to thank Michael Troup and Dinis Gökaydin for their helpful comments on the manuscript, as well as past and present members of the van Swinderen lab for discussions on the subject of active sleep in animals.

REFERENCES

Anderson, D. J. (2016). Circuit modules linking internal states and social behaviour in flies and mice. *Nat. Rev. Neurosci.* 17, 692–704. doi: 10.1038/nrn.2016.125

Andrillon, T., and Kouider, S. (2020). The vigilant sleeper: neural mechanisms of sensory (de)coupling during sleep. *Curr. Opin. Physiol.* 15, 47–59. doi: 10.1016/j.cophys.2019.12.002

Andrillon, T., Nir, Y., Cirelli, C., Tononi, G., and Fried, I. (2015). Single-neuron activity and eye movements during human REM sleep and awake vision. *Nat. Commun.* 6:ncomms8884. doi: 10.1038/ncomms8884

Arthaud, S., Varin, C., Gay, N., Libourel, P.-A., Chauveau, F., Fort, P., et al. (2015). Paradoxical (REM) sleep deprivation in mice using the small-platforms-over-water method: polysomnographic analyses and melanin-concentrating hormone and hypocretin/orexin neuronal activation before, during and after deprivation. *J. Sleep Res.* 24, 309–319. doi: 10.1111/jsr.12269

Aserinsky, E., and Kleitman, N. (1955). Two Types of Ocular Motility Occurring in Sleep. *J. Appl. Physiol.* 8, 1–10. doi: 10.1152/jappl.1955.8.1.1

Aserinsky, E., and Kleitman, N. (1953). Regularly Occurring Periods of Eye Motility, and Concomitant Phenomena, during Sleep. *Science* 118, 273–274.

Atienza, M., Cantero, J. L., and Gómez, C. M. (2000). Decay time of the auditory sensory memory trace during wakefulness and REM sleep. *Psychophysiology* 37, 485–493. doi: 10.1111/1469-8986.3740485

Ayala-Guerrero, F., and Mexicano, G. (2008). Sleep and wakefulness in the green iguanid lizard (*Iguana iguana*). *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* 151, 305–312. doi: 10.1016/j.cbpa.2007.03.027

Barbera, J. (2008). Sleep and dreaming in Greek and Roman philosophy. *Sleep Med.* 9, 906–910. doi: 10.1016/j.sleep.2007.10.010

Barbosa, L. S., and Kouider, S. (2018). Prior Expectation Modulates Repetition Suppression without Perceptual Awareness. *Sci. Rep.* 8:5055. doi: 10.1038/s41598-018-23467-3

Barrett, L. F., Quigley, K. S., and Hamilton, P. (2016). An active inference theory of allostasis and interoception in depression. *Philosoph. Transact. R. Soc. B Biol. Sci.* 371:20160011. doi: 10.1098/rstb.2016.0011

Barron, A. B., and Klein, C. (2016). What insects can tell us about the origins of consciousness. *Proc. Natl. Acad. Sci. U S A.* 113, 4900–4908. doi: 10.1073/pnas.1520084113

Barsky, M. M., Tucker, M. A., and Stickgold, R. (2015). REM Sleep Enhancement of Probabilistic Classification Learning is Sensitive to Subsequent Interference. *Neurobiol. Learn. Mem.* 122, 63–68. doi: 10.1016/j.nlm.2015.02.015

Barto, A., Mirolli, M., and Baldassarre, G. (2013). Novelty or Surprise? *Front. Psychol.* 4:907. doi: 10.3389/fpsyg.2013.00907

Bastuji, H., Perrin, F., and Garcia-Larrea, L. (2002). Semantic analysis of auditory input during sleep: studies with event related potentials. *Int. J. Psychophysiol. Event Rel. Potent. Measure Informat. Proces. During Sleep* 46, 243–255. doi: 10.1016/S0167-8760(02)00116-2

Bateson, M., Desire, S., Gartside, S. E., and Wright, G. A. (2011). Agitated Honeybees Exhibit Pessimistic Cognitive Biases. *Curr. Biol.* 21, 1070–1073. doi: 10.1016/j.cub.2011.05.017

Bedont, J. L., Toda, H., Shi, M., Park, C. H., Quake, C., Stein, C., et al. (2021). Short and long sleeping mutants reveal links between sleep and macroautophagy. *eLife* 10:e64140. doi: 10.7554/eLife.64140

Bekinschtein, T. A., Dehaene, S., Rohaut, B., Tadel, F., Cohen, L., and Naccache, L. (2009). Neural signature of the conscious processing of auditory regularities. *PNAS* 106, 1672–1677. doi: 10.1073/pnas.0809667106

Berger, M., and Riemann, D. (1993). REM sleep in depression—an overview. *J. Sleep Res.* 2, 211–223. doi: 10.1111/j.1365-2869.1993.tb00092.x

Bhatia, S., Mellers, B., and Walasek, L. (2019). Affective responses to uncertain real-world outcomes: Sentiment change on Twitter. *PLoS One* 14:e0212489. doi: 10.1371/journal.pone.0212489

Blagrove, M. (1993). Dreams as the reflection of our waking concerns and abilities: A critique of the problem-solving paradigm in dream research. *Dreaming* 2:205. doi: 10.1037/h0094361

Blake, H., and Gerard, R. W. (1937). Brain potentials during sleep. *Am. J. Physiol. Legacy Cont.* 119, 692–703. doi: 10.1152/ajplegacy.1937.119.4.692

Blakemore, S.-J., Wolpert, D., and Frith, C. (2000). Why can't you tickle yourself? *NeuroReport* 11:R11.

Bowers, M. B., Hartmann, E. L., and Freedman, D. X. (1966). Sleep Deprivation and Brain Acetylcholine. *Science* 153, 1416–1417.

Boyce, R., Glasgow, S. D., Williams, S., and Adamantidis, A. (2016). Causal evidence for the role of REM sleep theta rhythm in contextual memory consolidation. *Science* 352, 812–816. doi: 10.1126/science.aad5252

Braun, A. R., Balkin, T. J., Wesenten, N. J., Carson, R. E., Varga, M., Baldwin, P., et al. (1997). Regional cerebral blood flow throughout the sleep-wake cycle. An H2(15)O PET study. *Brain* 120(Pt 7), 1173–1197. doi: 10.1093/brain/120.7.1173

Cai, D. J., Mednick, S. A., Harrison, E. M., Kanady, J. C., and Mednick, S. C. (2009). REM, not incubation, improves creativity by priming associative networks. *PNAS* 106, 10130–10134. doi: 10.1073/pnas.0900271106

Campbell, S. S., and Tobler, I. (1984). Animal sleep: A review of sleep duration across phylogeny. *Neurosci. Biobehav. Rev.* 8, 269–300. doi: 10.1016/0149-7634(84)90054-X

Cao, J., Herman, A. B., West, G. B., Poe, G., and Savage, V. M. (2020). Unraveling why we sleep: Quantitative analysis reveals abrupt transition from neural reorganization to repair in early development. *Sci. Adv.* 6:eaba0398. doi: 10.1126/sciadv.aba0398

Card, G., and Dickinson, M. H. (2008). Visually Mediated Motor Planning in the Escape Response of Drosophila. *Curr. Biol.* 18, 1300–1307. doi: 10.1016/j.cub.2008.07.094

Carmody, D. P., and Lewis, M. (2006). Brain activation when hearing one's own and others' names. *Brain Res.* 1116, 153–158. doi: 10.1016/j.brainres.2006.07.121

Cartwright, R., Young, M. A., Mercer, P., and Bears, M. (1998). Role of REM sleep and dream variables in the prediction of remission from depression. *Psychiatry Res.* 80, 249–255. doi: 10.1016/S0165-1781(98)00071-7

Chen, T.-W., Wardill, T. J., Sun, Y., Pulver, S. R., Renninger, S. L., Baohan, A., et al. (2013). Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* 499, 295–300.

Cirelli, C. (2009). The genetic and molecular regulation of sleep: from fruit flies to humans. *Nat. Rev. Neurosci.* 10, 549–560. doi: 10.1038/nrn2683

Cirelli, C., Gutierrez, C. M., and Tononi, G. (2004). Extensive and Divergent Effects of Sleep and Wakefulness on Brain Gene Expression. *Neuron* 41, 35–43. doi: 10.1016/S0896-6273(03)00814-6

Clemes, S., and Dement, W. (1967). Effect of REM Sleep Deprivation on Psychological Functioning. *J. Nerv. Ment. Dis.* 144, 485–491.

Coons, S., and Guilleminault, C. (1982). Development of Sleep-Wake Patterns and Non-rapid Eye Movement Sleep Stages during the First Six Months of Life in Normal Infants. *Pediatrics* 69, 793–798.

Cote, K. A., Etienne, L., and Campbell, K. B. (2001). Neurophysiological Evidence for the Detection of External Stimuli During Sleep. *Sleep* 24, 1–13. doi: 10.1093/sleep/24.7.1

Crunelli, V., and Hughes, S. W. (2010). The slow (<1 Hz) rhythm of non-REM sleep: a dialogue between three cardinal oscillators. *Nat. Neurosci.* 13, 9–17. doi: 10.1038/nn.2445

D'Albis, T., Jaramillo, J., Sprekeler, H., and Kempter, R. (2015). Inheritance of Hippocampal Place Fields Through Hebbian Learning: Effects of Theta Modulation and Phase Precession on Structure Formation. *Neural Computat.* 27, 1624–1672. doi: 10.1162/NECO_a_00752

de Bivort, B. L., and van Swinderen, B. (2016). Evidence for selective attention in the insect brain. *Curr. Opin. Insect Sci. Pests Resist. Behav. Ecol.* 15, 9–15. doi: 10.1016/j.cois.2016.02.007

De Brigard, F., and Prinz, J. (2010). Attention and consciousness. *WIREs Cognit. Sci.* 1, 51–59. doi: 10.1002/wcs.27

de Vivo, L., Bellesi, M., Marshall, W., Bushong, E. A., Ellisman, M. H., Tononi, G., et al. (2017). Ultrastructural Evidence for Synaptic Scaling Across the Wake/sleep Cycle. *Science* 355, 507–510. doi: 10.1126/science.aah5982

Dement, W. (1960). The Effect of Dream Deprivation. *Science* 131, 1705–1707.

Diaz-Guerrero, R., Gottlieb, J. S., and Knott, J. R. (1946). The Sleep of Patients with Manic-Depressive Psychosis, Depressive Type: An Electroencephalographic Study. *Psychosomat. Med.* 8:399.

Dijk, D. J., Brunner, D. P., and Borbely, A. A. (1990). Time course of EEG power density during long sleep in humans. *Am. J. Physiol. Regulat. Integrat. Comparat. Physiol.* 258, R650–R661. doi: 10.1152/ajpregu.1990.258.3.R650

Dill, M., and Heisenberg, M. (1995). Visual Pattern Memory without Shape Recognition. *Philosop. Transact. Biol. Sci.* 349, 143–152.

Dubowy, C., and Sehgal, A. (2017). Circadian Rhythms and Sleep in Drosophila melanogaster. *Genetics* 205, 1373–1397. doi: 10.1534/genetics.115.185157

Dunlop, R., and Laming, P. (2005). Mechanoreceptive and Nociceptive Responses in the Central Nervous System of Goldfish (Carassius auratus) and Trout (Oncorhynchus mykiss). *J. Pain* 6, 561–568. doi: 10.1016/j.jpain.2005.02.010

Ermis, U., Krakow, K., and Voss, U. (2010). Arousal thresholds during human tonic and phasic REM sleep. *J. Sleep Res.* 19, 400–406. doi: 10.1111/j.1365-2869.2010.00831.x

Feather, N. T. (1967). Valence of outcome and expectation of success in relation to task difficulty and perceived locus of control. *J. Personal. Soc. Psychol.* 7, 372–386. doi: 10.1037/h0025184

Feldman, H., and Friston, K. (2010). Attention, Uncertainty, and Free-Energy. *Front. Hum. Neurosci.* 4:00215. doi: 10.3389/fnhum.2010.00215

Feuerriegel, D., Keage, H. A. D., Rossion, B., and Quek, G. L. (2018). Immediate stimulus repetition abolishes stimulus expectation and surprise effects in fast periodic visual oddball designs. *Biol. Psychol.* 138, 110–125. doi: 10.1016/j.biopsycho.2018.09.002

Finn, J. K., Tregenza, T., and Norman, M. D. (2009). Defensive tool use in a coconut-carrying octopus. *Curr. Biol.* 19, R1069–R1070. doi: 10.1016/j.cub.2009.10.052

Fischer, J., Iupac, and Ganellin, C. R. (2006). *Analogue-based Drug Discovery*. Hoboken, NJ: John Wiley & Sons.

Fisher, C., Gross, J., and Zuch, J. (1965). Cycle of Penile Erection Synchronous With Dreaming (REM) Sleep: Preliminary Report. *Arch. General Psychiatry* 12, 29–45. doi: 10.1001/archpsyc.1965.01720310031005

Fossat, P., Bacqué-Cazenave, J., De Deurwaerdère, P., Delbecque, J.-P., and Cattaert, D. (2014). Anxiety-like behavior in crayfish is controlled by serotonin. *Science* 344, 1293–1297.

Fotowat, H., Fayyazuddin, A., Bellen, H. J., and Gabbiani, F. (2009). A Novel Neuronal Pathway for Visually Guided Escape in Drosophila melanogaster. *J. Neurophysiol.* 102, 875–885. doi: 10.1152/jn.00073.2009

Frank, M. G., Waldrop, R. H., Dumoulin, M., Aton, S., and Boal, J. G. (2012). A Preliminary Analysis of Sleep-Like States in the Cuttlefish Sepia officinalis. *PLoS One* 7:e38125. doi: 10.1371/journal.pone.0038125

French, A. S., Geissmann, Q., Beckwith, E. J., and Gilestro, G. F. (2021). Sensory processing during sleep in Drosophila melanogaster. *Nature* 2021:03954–w. doi: 10.1038/s41586-021-03954-w

Friston, K. (2005). A theory of cortical responses. *Philosop. Transact. R. Soc. B Biol. Sci.* 360, 815–836. doi: 10.1098/rstb.2005.1622

Friston, K., Kilner, J., and Harrison, L. (2006). A free energy principle for the brain. *J. Physiol. Paris Theoret. Computat. Neurosci. Understand. Brain Funct.* 100, 70–87. doi: 10.1016/j.jphysparis.2006.10.001

Friston, K., Thornton, C., and Clark, A. (2012). Free-Energy Minimization and the Dark-Room Problem. *Front. Psychol.* 3:130. doi: 10.3389/fpsyg.2012.00130

Friston, K. J., Lin, M., Frith, C. D., Pezzulo, G., Hobson, J. A., and Ondobaka, S. (2017). Active Inference. *Curiosity Insight Neural Computat.* 29, 2633–2683. doi: 10.1162/neco_a_00999

Fujiwara, T., Cruz, T. L., Bohnslav, J. P., and Chiappe, M. E. (2017). A faithful internal representation of walking movements in the *Drosophila* visual system. *Nat. Neurosci.* 20, 72–81. doi: 10.1038/nn.4435

Fultz, N. E., Bonmassar, G., Setsompop, K., Stickgold, R. A., Rosen, B. R., Polimeni, J. R., et al. (2019). Coupled electrophysiological, hemodynamic, and cerebrospinal fluid oscillations in human sleep. *Science* 366, 628–631. doi: 10.1126/science.aax5440

Galbiati, A., Sforza, M., Fasiello, E., Casoni, F., Marrella, N., Leitner, C., et al. (2020). The association between emotional dysregulation and REM sleep features in insomnia disorder. *Brain Cognit.* 146:105642. doi: 10.1016/j.bandc.2020.105642

Gibson, W. T., Gonzalez, C. R., Fernandez, C., Ramasamy, L., Tabachnik, T., Du, R. R., et al. (2015). Behavioral Responses to a Repetitive Visual Threat Stimulus Express a Persistent State of Defensive Arousal in Drosophila. *Curr. Biol.* 25, 1401–1415. doi: 10.1016/j.cub.2015.03.058

Giedke, H., and Schwärzler, F. (2002). Therapeutic use of sleep deprivation in depression. *Sleep Med. Rev.* 6, 361–377. doi: 10.1053/smrv.2002.0235

Giuditta, A., Ambrosini, M. V., Montagnese, P., Mandile, P., Cotugno, M., Zucconi, G. G., et al. (1995). The sequential hypothesis of the function of sleep. *Behav. Brain Res.* 69, 157–166. doi: 10.1016/0166-4328(95)00012-I

Giurfa, M. (2007). Behavioral and neural analysis of associative learning in the honeybee: a taste from the magic well. *J. Comp. Physiol. A* 193, 801–824. doi: 10.1007/s00359-007-0235-9

Grabowska, M. J., Jeans, R., Steeves, J., and van Swinderen, B. (2020). Oscillations in the central brain of *Drosophila* are phase locked to attended visual features. *Proc. Natl. Acad. Sci. U S A.* 117, 29925–29936. doi: 10.1073/pnas.2010749117

Grabowska, M. J., Steeves, J., Alpay, J., Van De Poll, M., Ertekin, D., and van Swinderen, B. (2018). Innate visual preferences and behavioral flexibility in *Drosophila*. *J. Exp. Biol.* 221:jeb185918. doi: 10.1242/jeb.185918

Green, J. D., and Arduini, A. A. (1954). Hippocampal electrical activity in arousal. *J. Neurophysiol.* 17, 533–557. doi: 10.1152/jn.1954.17.6.533

Guzmán-marín, R., Suntsova, N., Stewart, D. R., Gong, H., Szymusiak, R., and McGinty, D. (2003). Sleep deprivation reduces proliferation of cells in the dentate gyrus of the hippocampus in rats. *J. Physiol.* 549, 563–571. doi: 10.1113/jphysiol.2003.041665

Harrington, M. O., Johnson, J. M., Croom, H. E., Pennington, K., and Durrant, S. J. (2018). The influence of REM sleep and SWS on emotional memory consolidation in participants reporting depressive symptoms. *Cortex* 99, 281–295. doi: 10.1016/j.cortex.2017.12.004

Heisenberg, M., and Wolf, R. (1984). *Vision in Drosophila: Genetics of Microbehavior, Studies of Brain Function*. Berlin: Springer-Verlag. doi: 10.1007/978-3-642-69936-8

Helmholtz, H. (1860). *Handbuch der Physiologischen Optik,* Vol. 3, trans. J. P. C. Southall. New York, NY: Dover.

Henane, R., Buguet, A., Roussel, B., and Bittel, J. (1977). Variations in evaporation and body temperatures during sleep in man. *J. Appl. Physiol.* 42, 50–55. doi: 10.1152/jappl.1977.42.1.50

Hendricks, J. C., Finn, S. M., Panckeri, K. A., Chavkin, J., Williams, J. A., Sehgal, A., et al. (2000). Rest in Drosophila Is a Sleep-like State. *Neuron* 25, 129–138. doi: 10.1016/S0896-6273(00)80877-6

Herman, M. D., Denlinger, S. L., Patarca, R., Katz, L., and Hobson, J. A. (1991). Developmental phases of sleep and motor behaviour in a cat mother-infant system: A time-lapse video approach. *Canad. J. Psychol.* 45:101. doi: 10.1037/h0084278

Hill, A. J., Mansfield, R., Lopez, J. M., Raizen, D. M., and Van Buskirk, C. (2014). Cellular Stress Induces a Protective Sleep-like State in *C. elegans*. *Curr. Biol.* 24, 2399–2405. doi: 10.1016/j.cub.2014.08.040

Hinton, G. E., Dayan, P., Frey, B. J., and Neal, R. M. (1995). The "Wake-Sleep" Algorithm for Unsupervised Neural Networks. *Science* 268, 1158–1161.

Hobson, J. A. (2009a). REM sleep and dreaming: towards a theory of protoconsciousness. *Nat. Rev. Neurosci.* 10, 803–813. doi: 10.1038/nrn2716

Hobson, J. A. (2009b). The AIM Model of Dreaming, Sleeping, and Waking Consciousness. *Encyclop. Neurosci.* 2009, 963–970. doi: 10.1016/B978-008045046-9.00042-5

Hobson, J. A., and Friston, K. J. (2012). Waking and dreaming consciousness: Neurobiological and functional considerations. *Progress Neurobiol.* 98, 82–98. doi: 10.1016/j.pneurobio.2012.05.003

Hobson, J. A., Gott, J. A., and Friston, K. J. (2021). Minds and Brains, Sleep and Psychiatry. *Psychiatric Res. Clin. Pract.* 3, 12–28. doi: 10.1176/appi.prcp.20200023

Hobson, J. A., Hong, C. C.-H., and Friston, K. J. (2014). Virtual reality and consciousness inference in dreaming. *Front. Psychol.* 5:1133. doi: 10.3389/fpsyg.2014.01133

Hoel, E. (2021). The overfitted brain: Dreams evolved to assist generalization. *Patterns* 2:100244. doi: 10.1016/j.patter.2021.100244

Hong, C. C.-H., Harris, J. C., Pearlson, G. D., Kim, J.-S., Calhoun, V. D., Fallon, J. H., et al. (2009). fMRI evidence for multisensory recruitment associated with rapid eye movements during sleep. *Hum. Brain Mapp.* 30, 1705–1722. doi: 10.1002/hbm.20635

Hoopfer, E. D. (2016). Neural control of aggression in Drosophila. *Curr. Opin. Neurobiol. Neurobiol. Sex* 38, 109–118. doi: 10.1016/j.conb.2016.04.007

Hu, W., Peng, Y., Sun, J., Zhang, F., Zhang, X., Wang, L., et al. (2018). Fan-Shaped Body Neurons in the Drosophila Brain Regulate Both Innate and Conditioned Nociceptive Avoidance. *Cell Rep.* 24, 1573–1584. doi: 10.1016/j.celrep.2018.07.028

Huber, R., Felice Ghilardi, M., Massimini, M., and Tononi, G. (2004). Local sleep and learning. *Nature* 430, 78–81. doi: 10.1038/nature02663

Huettel, S. A., and McCarthy, G. (2004). What is odd in the oddball task?: Prefrontal cortex is activated by dynamic changes in response strategy. *Neuropsychologia* 42, 379–386. doi: 10.1016/j.neuropsychologia.2003.07.009

Ibáñez, A. M., Martín, R. S., Hurtado, E., and López, V. (2009). ERPs studies of cognitive processing during sleep. *Int. J. Psychol.* 44, 290–304. doi: 10.1080/00207590802194234

Iglesias, T. L., Boal, J. G., Frank, M. G., Zeil, J., and Hanlon, R. T. (2019). Cyclic nature of the REM sleep-like state in the cuttlefish Sepia officinalis. *J. Exp. Biol.* 222:jeb174862. doi: 10.1242/jeb.174862

Ishimori, K. (1909). True cause of sleep: a hypnogenic substance as evidenced in the brain of sleep-deprived animals. *Tokyo Igakkai Zasshi* 23, 429–457.

Ishimoto, H., Lark, A. R. S., and Kitamoto, T. (2012). Factors that Differentially Affect Daytime and Nighttime Sleep in Drosophila melanogaster. *Front. Neurol.* 3:00024. doi: 10.3389/fneur.2012.00024

Jaramillo, J., and Kempter, R. (2017). Phase precession: a neural code underlying episodic memory? *Curr. Opin. Neurobiol.* 43, 130–138. doi: 10.1016/j.conb.2017.02.006

Jessen, N. A., Munk, A. S. F., Lundgaard, I., and Nedergaard, M. (2015). The Glymphatic System – A Beginner's Guide. *Neurochem. Res.* 40, 2583–2599. doi: 10.1007/s11064-015-1581-6

Jouvet, M. (1961). "Telencephalic and Rhombencephalic Sleep in the Cat," in *Ciba Foundation Symposium - The Nature of Sleep* (Hoboken, NJ: John Wiley & Sons, Ltd). doi: 10.1002/9780470719220.ch9

Jouvet-Mounier, D., Astic, L., and Lacote, D. (1969). Ontogenesis of the states of sleep in rat, cat, and guinea pig during the first postnatal month. *Dev. Psychobiol.* 2, 216–239. doi: 10.1002/dev.420020407

Kaiser, W. (1988). Busy bees need rest, too. *J. Comp. Physiol.* 163, 565–584. doi: 10.1007/BF00603841

Kaiser, W., and Steiner-Kaiser, J. (1983). Neuronal correlates of sleep, wakefulness and arousal in a diurnal insect. *Nature* 301, 707–709. doi: 10.1038/301707a0

Kanai, R., Komura, Y., Shipp, S., and Friston, K. (2015). Cerebral hierarchies: predictive processing, precision and the pulvinar. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370:0169. doi: 10.1098/rstb.2014.0169

Kanaya, H. J., Park, S., Kim, J., Kusumi, J., Krenenou, S., Sawatari, E., et al. (2020). A sleep-like state in Hydra unravels conserved sleep mechanisms during the evolutionary development of the central nervous system. *Sci. Adv.* 6:eabb9415. doi: 10.1126/sciadv.abb9415

Karni, A., Tanne, D., Rubenstein, B. S., Askenasy, J. J. M., and Sagi, D. (1994). Dependence on REM Sleep of Overnight Improvement of a Perceptual Skill. *Science* 265, 679–682.

Kaufman, L. S., and Morrison, A. R. (1981). Spontaneous and elicited PGO spikes in rats. *Brain Res.* 214, 61–72. doi: 10.1016/0006-8993(81)90438-8

Kayser, M. S., and Biron, D. (2016). Sleep and Development in Genetically Tractable Model Organisms. *Genetics* 203, 21–33. doi: 10.1534/genetics.116.189589

Kim, A. J., Fitzgerald, J. K., and Maimon, G. (2015). Cellular evidence for efference copy in *Drosophila* visuomotor processing. *Nat. Neurosci.* 18, 1247–1255. doi: 10.1038/nn.4083

Kirszenblat, L., and van Swinderen, B. (2015). The Yin and Yang of Sleep and Attention. *Trends Neurosci.* 38, 776–786. doi: 10.1016/j.tins.2015.10.001

Klein, B. A., Stiegler, M., Klein, A., and Tautz, J. (2014). Mapping Sleeping Bees within Their Nest: Spatial and Temporal Analysis of Worker Honey Bee Sleep. *PLoS One* 9:e102316. doi: 10.1371/journal.pone.0102316

Koelsch, S., Heinke, W., Sammler, D., and Olthoff, D. (2006). Auditory processing during deep propofol sedation and recovery from unconsciousness. *Clin. Neurophysiol.* 117, 1746–1759. doi: 10.1016/j.clinph.2006.05.009

Kravitz, E. A., and Huber, R. (2003). Aggression in invertebrates. *Curr. Opin. Neurobiol.* 13, 736–743. doi: 10.1016/j.conb.2003.10.003

Labroo, A. A., Mukhopadhyay, A., and Dong, P. (2014). Not always the best medicine: Why frequent smiling can reduce wellbeing. *J. Exp. Soc. Psychol.* 53, 156–162. doi: 10.1016/j.jesp.2014.03.001

Lee, A. K., and Wilson, M. A. (2002). Memory of Sequential Experience in the Hippocampus during Slow Wave Sleep. *Neuron* 36, 1183–1194. doi: 10.1016/S0896-6273(02)01096-6

Legendre, R., and Piéron, H. (1913). Recherches sur le besoin de sommeil consecutif à une veille prolongée. *Z. Allgem. Physiol.* 14, 235–262.

Lesku, J. A., Aulsebrook, A. E., Kelly, M. L., and Tisdale, R. K. (2019). "Chapter 20 - Evolution of Sleep and Adaptive Sleeplessness," in *Handbook of Behavioral Neuroscience, Handbook of Sleep Research*, ed. H. C. Dringenberg (Amsterdam: Elsevier), 299–316. doi: 10.1016/B978-0-12-813743-7.00020-7

Leung, L. C., Wang, G. X., Madelaine, R., Skariah, G., Kawakami, K., Deisseroth, K., et al. (2019). Neural signatures of sleep in zebrafish. *Nature* 571:198. doi: 10.1038/s41586-019-1336-7

Libourel, P.-A., and Herrel, A. (2016). Sleep in amphibians and reptiles: a review and a preliminary analysis of evolutionary patterns. *Biol. Rev.* 91, 833–866. doi: 10.1111/brv.12197

Liu, G., Seiler, H., Wen, A., Zars, T., Ito, K., Wolf, R., et al. (2006). Distinct memory traces for two visual features in the Drosophila brain. *Nature* 439, 551–556. doi: 10.1038/nature04381

Llewellyn, S. (2016). Dream to predict? REM dreaming as prospective coding. *Front. Psychol.* 6:01961. doi: 10.3389/fpsyg.2015.01961

Loomis, A. L., Harvey, E. N., and Hobart, G. A. (1937). Cerebral states during sleep, as studied by human brain potentials. *J. Exp. Psychol.* 21:127. doi: 10.1037/h0057431

Lu, J., Sherman, D., Devor, M., and Saper, C. B. (2006). A putative flip–flop switch for control of REM sleep. *Nature* 441, 589–594. doi: 10.1038/nature04767

Lutz, N. D., Wolf, I., Hübner, S., Born, J., and Rauss, K. (2018). Sleep Strengthens Predictive Sequence Coding. *J. Neurosci.* 38, 8989–9000. doi: 10.1523/JNEUROSCI.1352-18.2018

Ly, S., Pack, A. I., and Naidoo, N. (2018). The neurobiological basis of sleep: Insights from Drosophila. *Neurosci. Biobehav. Rev.* 87, 67–86. doi: 10.1016/j.neubiorev.2018.01.015

Malafeev, A., Laptev, D., Bauer, S., Omlin, X., Wierzbicka, A., Wichniak, A., et al. (2018). Automatic Human Sleep Stage Scoring Using Deep Neural Networks. *Front. Neurosci.* 12:00781. doi: 10.3389/fnins.2018.00781

McDonald, D. G., Schicht, W. W., Frazier, R. E., Shallenberger, H. D., and Edwards, D. J. (1975). Studies of Information Processing in Sleep. *Psychophysiology* 12, 624–629. doi: 10.1111/j.1469-8986.1975.tb00059.x

Meddis, R. (1975). On the function of sleep. *Anim. Behav.* 23, 676–691. doi: 10.1016/0003-3472(75)90144-X

Medeiros, S. L., de, S., Paiva, M. M. M., de, Lopes, P. H., Blanco, W., et al. (2021). Cyclic alternation of quiet and active sleep states in the octopus. *iScience* 2021:102223. doi: 10.1016/j.isci.2021.102223

Mendoza-Angeles, K., Cabrera, A., Hernández-Falcón, J., and Ramón, F. (2007). Slow waves during sleep in crayfish: A time–frequency analysis. *J. Neurosci. Methods* 162, 264–271. doi: 10.1016/j.jneumeth.2007.01.025

Mendoza-Angeles, K., Hernández-Falcón, J., and Ramón, F. (2010). Slow waves during sleep in crayfish. Origin and spread. *J. Exp. Biol.* 213, 2154–2164. doi: 10.1242/jeb.038240

Miyazaki, S., Liu, C.-Y., and Hayashi, Y. (2017). Sleep in vertebrate and invertebrate animals, and insights into the function and evolution of sleep. *Neurosci. Res.* 118, 3–12. doi: 10.1016/j.neures.2017.04.017

Morden, B., Conner, R., Mitchell, G., Dement, W., and Levine, S. (1968). Effects of rapid eye movement (REM) sleep deprivation on shock-induced fighting☆☆☆. *Physiol. Behav.* 3, 425–432. doi: 10.1016/0031-9384(68)90073-5

Näätänen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behav. Brain Sci.* 13, 201–233. doi: 10.1017/S0140525X00078407

Nath, R. D., Bedbrook, C. N., Abrams, M. J., Basinger, T., Bois, J. S., Prober, D. A., et al. (2017). The Jellyfish Cassiopea Exhibits a Sleep-like State. *Curr. Biol.* 27, 2984.e–2990.e. doi: 10.1016/j.cub.2017.08.014

Netchiporouk, L., Shram, N., Salvert, D., and Cespuglio, R. (2001). Brain extracellular glucose assessed by voltammetry throughout the rat sleep–wake cycle. *Eur. J. Neurosci.* 13, 1429–1434. doi: 10.1046/j.0953-816x.2001.01503.x

Nichols, A. L. A., Eichler, T., Latham, R., and Zimmer, M. (2017). A global brain state underlies C. elegans sleep behavior. *Science* 356:aam6851. doi: 10.1126/science.aam6851

Nielsen, T. A., Deslauriers, D., and Baylor, G. W. (1992). Emotions in dream and waking event reports. *Dreaming* 1:287. doi: 10.1037/h0094340

Nishida, M., Pearsall, J., Buckner, R. L., and Walker, M. P. (2009). REM Sleep, Prefrontal Theta, and the Consolidation of Human Emotional Memory. *Cereb. Cortex* 19, 1158–1166. doi: 10.1093/cercor/bhn155

Niwa, Y., Kanda, G. N., Yamada, R. G., Shi, S., Sunagawa, G. A., Ukai-Tadenuma, M., et al. (2018). Muscarinic Acetylcholine Receptors Chrm1 and Chrm3 Are Essential for REM Sleep. *Cell Rep.* 24, 2231.e–2247.e. doi: 10.1016/j.celrep.2018.07.082

Ogawa, K., and Otani, E. (2014). Role of REM sleep in the emotional brain regulation for social pain. *Int. J. Psychophysiol.* 94:173. doi: 10.1016/j.ijpsycho.2014.08.741

Okawa, M., Mishima, K., Hishikawa, Y., Hozumi, S., Hori, H., and Takahashi, K. (1991). Circadian Rhythm Disorders in Sleep-Waking and Body Temperature in Elderly Patients with Dementia and Their Treatment. *Sleep* 14, 478–485. doi: 10.1093/sleep/14.6.478

O'Keefe, J., and Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Res.* 34, 171–175. doi: 10.1016/0006-8993(71)90358-1

Oswald, I., Taylor, A. M., and Treisman, M. (1960). Discriminative responses to stimulation during human sleep. *Brain* 83, 440–453. doi: 10.1093/brain/83.3.440

Parmeggiani, P. (1990). Thermoregulation During Sleep in Mammals. *Physiology* 5, 208–212. doi: 10.1152/physiologyonline.1990.5.5.208

Peirano, P., Algarín, C., and Uauy, R. (2003). Sleep-wake states and their regulatory mechanisms throughout early human development. *J. Pediatr.* 143, 70–79. doi: 10.1067/S0022-3476(03)00404-9

Picton, T. W. (1992). The P300 Wave of the Human Event-Related Potential. *J. Clin. Neurophysiol.* 9, 456–479.

Plana-Ripoll, O., Pedersen, C. B., Holtz, Y., Benros, M. E., Dalsgaard, S., de Jonge, P., et al. (2019). Exploring Comorbidity Within Mental Disorders Among a Danish National Population. *JAMA Psychiatry* 76, 259–270. doi: 10.1001/jamapsychiatry.2018.3658

Prober, D. A., Rihel, J., Onah, A. A., Sung, R.-J., and Schier, A. F. (2006). Hypocretin/Orexin Overexpression Induces An Insomnia-Like Phenotype in Zebrafish. *J. Neurosci.* 26, 13400–13410. doi: 10.1523/JNEUROSCI.4332-06.2006

Queiroz, C. M., Tiba, P. A., Moreira, K. M., Guidine, P. A. M., Rezende, G. H. S., Moraes, M. F. D., et al. (2013). Sleep pattern and learning in knockdown mice with reduced cholinergic neurotransmission. *Braz. J. Med. Biol. Res.* 46, 844–854. doi: 10.1590/1414-431X20133102

Raccuglia, D., Huang, S., Ender, A., Heim, M.-M., Laber, D., Suárez-Grimalt, R., et al. (2019). Network-Specific Synchronization of Electrical Slow-Wave Oscillations Regulates Sleep Drive in Drosophila. *Curr. Biol.* 29, 3611.e–3621.e. doi: 10.1016/j.cub.2019.08.070

Raizen, D. M., Zimmerman, J. E., Maycock, M. H., Ta, U. D., You, Y., Sundaram, M. V., et al. (2008). Lethargus is a Caenorhabditis elegans sleep-like state. *Nature* 451, 569–572. doi: 10.1038/nature06535

Ramón, F., Hernández-Falcón, J., Nguyen, B., and Bullock, T. H. (2004). Slow Wave Sleep in Crayfish. *Proc. Natl. Acad. Sci. U S A.* 101, 11857–11861.

Rao, R. P. N., and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87. doi: 10.1038/4580

Reyes, R. B., Hill, S. Y., and Kupfer, D. J. (1983). Effects of acute doses of zimelidine on REM sleep in rats. *Psychopharmacology* 80, 214–216. doi: 10.1007/BF00436155

Riemann, D., Velthaus, S., Laubenthal, S., Müller, W. E., and Berger, M. (1990). REM-Suppressing Effects of Amitriptyline and Amitriptyline-N-Oxide After Acute Medication in Healthy Volunteers: Results of Two Uncontrolled Pilot Trials. *Pharmacopsychiatry* 23, 253–258. doi: 10.1055/s-2007-1014515

Roffwarg, H. P., Muzio, J. N., and Dement, W. C. (1966). Ontogenetic Development of the Human Sleep-Dream Cycle. *Science* 152, 604–619. doi: 10.1126/science.152.3722.604

Rosales-Lagarde, A., Armony, J. L., del Río-Portilla, Y., Trejo-Martínez, D., Conde, R., and Corsi-Cabrera, M. (2012). Enhanced emotional reactivity after selective REM sleep deprivation in humans: an fMRI study. *Front. Behav. Neurosci.* 6:00025. doi: 10.3389/fnbeh.2012.00025

Sabri, M., and Campbell, K. B. (2005). Is the failure to detect stimulus deviance during sleep due to a rapid fading of sensory memory or a degradation of stimulus encoding? *J. Sleep Res.* 14, 113–122. doi: 10.1111/j.1365-2869.2005.00446.x

Sareen, P., McCurdy, L. Y., and Nitabach, M. N. (2020). A neuronal ensemble encoding adaptive choice during sensory conflict in Drosophila. *Nat. Commun.* 12:4131.

Sareen, P., Wolf, R., and Heisenberg, M. (2011). Attracting the attention of a fly. *PNAS* 108, 7230–7235. doi: 10.1073/pnas.1102522108

Sassin, J. F., Parker, D. C., Mace, J. W., Gotlin, R. W., Johnson, L. C., and Rossman, L. G. (1969). Human Growth Hormone Release: Relation to Slow-Wave Sleep and Sleep-Waking Cycles. *Science* 165, 513–515.

Sauer, S., Kinkelin, M., Herrmann, E., and Kaiser, W. (2003). The dynamics of sleep-like behaviour in honey bees. *J. Comp. Physiol. A* 189, 599–607. doi: 10.1007/s00359-003-0436-9

Schachter, S., and Singer, J. (1963). Cognitive, social, and physiological determinants of emotional state. *Psychol. Rev.* 69:379. doi: 10.1037/h0046234

Scheel, D., Godfrey-Smith, P., and Lawrence, M. (2016). Signal Use by Octopuses in Agonistic Interactions. *Curr. Biol.* 26, 377–382. doi: 10.1016/j.cub.2015.12.033

Schwartenbeck, P., FitzGerald, T., Dolan, R., and Friston, K. (2013). Exploration, novelty, surprise, and free energy minimization. *Front. Psychol.* 4:710. doi: 10.3389/fpsyg.2013.00710

Sculthorpe, L. D., Ouellet, D. R., and Campbell, K. B. (2009). MMN elicitation during natural sleep to violations of an auditory pattern. *Brain Res.* 1290, 52–62. doi: 10.1016/j.brainres.2009.06.013

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends Cognit. Sci.* 17, 565–573. doi: 10.1016/j.tics.2013.09.007

Seymour, J. E., Carrette, T. J., and Sutherland, P. A. (2004). Do box jellyfish sleep at night? *Med. J. Austral.* 181, 707–707. doi: 10.5694/j.1326-5377.2004.tb06529.x

Shaw, P. J., Cirelli, C., Greenspan, R. J., and Tononi, G. (2000). Correlates of Sleep and Waking in Drosophila melanogaster. *Science* 287, 1834–1837. doi: 10.1126/science.287.5459.1834

Shein-Idelson, M., Ondracek, J. M., Liaw, H.-P., Reiter, S., and Laurent, G. (2016). Slow waves, sharp waves, ripples, and REM in sleeping dragons. *Science* 352, 590–595. doi: 10.1126/science.aaf3621

Shipp, S. (2016). Neural Elements for Predictive Coding. *Front. Psychol.* 7:01792. doi: 10.3389/fpsyg.2016.01792

Siegel, J. M. (2008). Do all animals sleep? *Trends Neurosci.* 31, 208–213. doi: 10.1016/j.tins.2008.02.001

Sippel, D., Schwabedal, J., Snyder, J. C., Oyanedel, C. N., Bernas, S. N., Garthe, A., et al. (2020). Disruption of NREM sleep and sleep-related spatial memory consolidation in mice lacking adult hippocampal neurogenesis. *Sci. Rep.* 10:16467. doi: 10.1038/s41598-020-72362-3

Skaggs, W. E., McNaughton, B. L., Wilson, M. A., and Barnes, C. A. (1996). Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus* 6, 149–172.

Smith, C. (1985). Sleep states and learning: A review of the animal literature. *Neurosci. Biobehav. Rev.* 9, 157–168. doi: 10.1016/0149-7634(85)90042-9

Solanki, N., Wolf, R., and Heisenberg, M. (2015). Central complex and mushroom bodies mediate novelty choice behavior in Drosophila. *J. Neurogenet.* 29, 30–37. doi: 10.3109/01677063.2014.1002661

Solvi, C., Baciadonna, L., and Chittka, L. (2016). Unexpected rewards induce dopamine-dependent positive emotion–like state changes in bumblebees. *Science* 353, 1529–1531. doi: 10.1126/science.aaf4454

Spoormaker, V. I., Schröter, M. S., Andrade, K. C., Dresler, M., Kiem, S. A., Goya-Maldonado, R., et al. (2012). Effects of rapid eye movement sleep deprivation on fear extinction recall and prediction error signaling. *Hum. Brain Mapp.* 33, 2362–2376. doi: 10.1002/hbm.21369

Stanhope, B. A., Jaggard, J. B., Gratton, M., Brown, E. B., and Keene, A. C. (2020). Sleep Regulates Glial Plasticity and Expression of the Engulfment Receptor Draper Following Neural Injury. *Curr. Biol.* 30, 1092.e–1101.e. doi: 10.1016/j.cub.2020.02.057

Sten, T. H., Li, R., Otopalik, A., and Ruta, V. (2021). Sexual arousal gates visual processing during Drosophila courtship. *Nature* 595, 549–553. doi: 10.1038/s41586-021-03714-w

Stock, A.-K., Wascher, E., and Beste, C. (2013). Differential Effects of Motor Efference Copies and Proprioceptive Information on Response Evaluation Processes. *PLoS One* 8:e62335. doi: 10.1371/journal.pone.0062335

Strauss, M., Sitt, J. D., King, J.-R., Elbaz, M., Azizi, L., Buiatti, M., et al. (2015). Disruption of hierarchical predictive coding during sleep. *PNAS* 112, E1353–E1362. doi: 10.1073/pnas.1501026112

Tainton-Heap, L. A. L., Kirszenblat, L. C., Notaras, E. T., Grabowska, M. J., Jeans, R., Feng, K., et al. (2021). A paradoxical kind of sleep in Drosophila melanogaster. *Curr. Biol.* 31, 578–590e6. doi: 10.1016/j.cub.2020.10.081

Tauber, E. S., Roffwarg, H. P., and Weitzman, E. D. (1966). Eye Movements and Electroencephalogram Activity during Sleep in Diurnal Lizards. *Nature* 212:1612. doi: 10.1038/2121612a0

Tesche, C. D., and Karhu, J. (2000). Theta oscillations index human hippocampal activation during a working memory task. *Proc. Natl. Acad. Sci. U. S. A.* 97, 919–924. doi: 10.1073/pnas.97.2.919

Tobler, I., and Neuner-Jehle, M. (1992). 24-h variation of vigilance in the cockroach Blaberus giganteus. *J. Sleep Res.* 1, 231–239. doi: 10.1111/j.1365-2869.1992.tb00044.x

Tononi, G., and Cirelli, C. (2014). Sleep and the Price of Plasticity: From Synaptic and Cellular Homeostasis to Memory Consolidation and Integration. *Neuron* 81, 12–34. doi: 10.1016/j.neuron.2013.12.025

Tononi, G., and Cirelli, C. (2006). Sleep function and synaptic homeostasis. *Sleep Med. Rev.* 10, 49–62. doi: 10.1016/j.smrv.2005.05.002

Troup, M., Yap, M. H., Rohrscheib, C., Grabowska, M. J., Ertekin, D., Randeniya, R., et al. (2018). Acute control of the sleep switch in Drosophila reveals a role for gap junctions in regulating behavioral responsiveness. *eLife* 7:37105. doi: 10.7554/eLife.37105

Ulanovsky, N., and Moss, C. F. (2007). Hippocampal cellular and network activity in freely moving echolocating bats. *Nat. Neurosci.* 10, 224–233. doi: 10.1038/nn1829

van Alphen, B., Semenza, E. R., Yap, M., Swinderen, B., van, and Allada, R. (2021). A deep sleep stage in Drosophila with a functional role in waste clearance. *Sci. Adv.* 7:eabc2999. doi: 10.1126/sciadv.abc2999

van Alphen, B., Yap, M. H. W., Kirszenblat, L., Kottler, B., and van Swinderen, B. (2013). A Dynamic Deep Sleep Stage in Drosophila. *J. Neurosci.* 33, 6917–6927. doi: 10.1523/JNEUROSCI.0061-13.2013

van Boxtel, J. J. A., Tsuchiya, N., and Koch, C. (2010). Consciousness and Attention: On Sufficiency and Necessity. *Front. Psychol.* 1:217. doi: 10.3389/fpsyg.2010.00217

van Swinderen, B. (2007). Attention-like processes in Drosophila require short-term memory genes. *Science* 315, 1590–1593. doi: 10.1126/science.1137931

van Swinderen, B. (2011). "Attention in Drosophila," in *International Review of Neurobiology, Recent Advances in the Use of Drosophila in Neurobiology and Neurodegeneration*, ed. N. Atkinson (Cambridge, MA: Academic Press), 51–85. doi: 10.1016/B978-0-12-387003-2.00003-3

van Swinderen, B., and Greenspan, R. J. (2003). Salience modulates 20–30 Hz brain activity in Drosophila. *Nat. Neurosci.* 6, 579–586. doi: 10.1038/nn1054

Villano, W. J., Otto, A. R., Ezie, C. E. C., Gillis, R., and Heller, A. S. (2020). Temporal dynamics of real-world emotion are more strongly linked to prediction error than outcome. *J. Exp. Psychol. General* 149:1755. doi: 10.1037/xge0000740

Vogel, G. W. (1968). REM Deprivation: III. Dreaming and Psychosis. *Arch. General Psychiatry* 18, 312–329. doi: 10.1001/archpsyc.1968.01740030056007

Vogel, G. W., Vogel, F., McAbee, R. S., and Thurmond, A. J. (1980). Improvement of Depression by REM Sleep Deprivation: New Findings and a Theory. *Arch. General Psychiatry* 37, 247–253. doi: 10.1001/archpsyc.1980.01780160017001

von Frisch, K. (1967). *The Dance Language and Orientation of Bees*, trans. ed. L. E. Chadwick. Cambridge, MA: The Belknap Press of Harvard University Press.

Vyazovskiy, V. V., and Harris, K. D. (2013). Sleep and the single neuron: the role of global slow oscillations in individual cell rest. *Nat. Rev. Neurosci.* 14, 443–451. doi: 10.1038/nrn3494

Wagner, U., Gais, S., Haider, H., Verleger, R., and Born, J. (2004). Sleep inspires insight. *Nature* 427, 352–355. doi: 10.1038/nature02223

Waters, F., Chiu, V., Atkinson, A., and Blom, J. D. (2018). Severe Sleep Deprivation Causes Hallucinations and a Gradual Progression Toward Psychosis With Increasing Time Awake. *Front. Psychiatry* 9:303. doi: 10.3389/fpsyt.2018.00303

Wehrle, R., Kaufmann, C., Wetter, T. C., Holsboer, F., Auer, D. P., Pollmächer, T., et al. (2007). Functional microstates within human REM sleep: first evidence from fMRI of a thalamocortical network specific for phasic REM periods: Thalamocortical network in phasic REM sleep. *Eur. J. Neurosci.* 25, 863–871. doi: 10.1111/j.1460-9568.2007.05314.x

Wiesner, C. D., Pulst, J., Krause, F., Elsner, M., Baving, L., Pedersen, A., et al. (2015). The effect of selective REM-sleep deprivation on the consolidation and affective evaluation of emotional memories. *Neurobiol. Learn. Mem. REM Sleep Mem.* 122, 131–141. doi: 10.1016/j.nlm.2015.02.008

Wiggin, T. D., Goodwin, P. R., Donelson, N. C., Liu, C., Trinh, K., Sanyal, S., et al. (2020). Covert sleep-related biological processes are revealed by probabilistic analysis in *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* 117, 10024–10034. doi: 10.1073/pnas.1917573117

Windt, J. M. (2021). How deep is the rift between conscious states in sleep and wakefulness? Spontaneous experience over the sleep–wake cycle. *Philosop. Transact. R. Soc. B Biol. Sci.* 376:20190696. doi: 10.1098/rstb.2019.0696

Windt, J. M. (2018). Predictive brains, dreaming selves, sleeping bodies: how the analysis of dream movement can inform a theory of self- and world-simulation in dreams. *Synthese* 195, 2577–2625. doi: 10.1007/s11229-017-1525-6

Xie, L., Kang, H., Xu, Q., Chen, M. J., Liao, Y., Thiyagarajan, M., et al. (2013). Sleep Drives Metabolite Clearance from the Adult Brain. *Science* 342, 373–377. doi: 10.1126/science.1241224

Xu, X., Yang, W., Tian, B., Sui, X., Chi, W., Rao, Y., et al. (2021). Quantitative investigation reveals distinct phases in Drosophila sleep. *Commun. Biol.* 4, 1–11. doi: 10.1038/s42003-021-01883-y

Yap, M. H. W., Grabowska, M. J., Rohrscheib, C., Jeans, R., Troup, M., Paulk, A. C., et al. (2017). Oscillatory brain activity in spontaneous and induced sleep stages in flies. *Nat. Commun.* 8:1815. doi: 10.1038/s41467-017-02024-y

Yetkin, S., Aydin, H., and Özgen, F. (2010). Polysomnography in patients with post-traumatic stress disorder. *Psychiatry Clin. Neurosci.* 64, 309–317. doi: 10.1111/j.1440-1819.2010.02084.x

Yokogawa, T., Marin, W., Faraco, J., Pézeron, G., Appelbaum, L., Zhang, J., et al. (2007). Characterization of Sleep in Zebrafish and Insomnia in Hypocretin Receptor Mutants. *PLoS Biol.* 5:e277. doi: 10.1371/journal.pbio.0050277

Young, J. Z. (1991). Light has many meanings for cephalopods. *Visual Neurosci.* 7, 1–12. doi: 10.1017/S0952523800010907

Zada, D., Bronshtein, I., Lerer-Goldshtein, T., Garini, Y., and Appelbaum, L. (2019). Sleep increases chromosome dynamics to enable reduction of accumulating DNA damage in single neurons. *Nat. Commun.* 10:895. doi: 10.1038/s41467-019-08806-w

Zarcone, V. P. Jr., Benson, K. L., and Berger, P. A. (1987). Abnormal Rapid Eye Movement Latencies in Schizophrenia. *Arch. General Psychiatry* 44, 45–48. doi: 10.1001/archpsyc.1987.01800130047007

Zhdanova, I. V., Wang, S. Y., Leclair, O. U., and Danilova, N. P. (2001). Melatonin promotes sleep-like state in zebrafish11Published on the World Wide Web on 6 April 2001. *Brain Res.* 903, 263–268. doi: 10.1016/S0006-8993(01)02444-1

Zupanc, G. K. H. (2006). Theodore H. Bullock (1915–2005). *Nature* 439, 280–280. doi: 10.1038/439280a

Zwaka, H., Bartels, R., Gora, J., Franck, V., Culo, A., Götsch, M., et al. (2015). Context Odor Presentation during Sleep Enhances Memory in Honeybees. *Curr. Biol.* 25, 2869–2874. doi: 10.1016/j.cub.2015.09.069

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Systems Neuroscience

ORIGINAL RESEARCH
published: 09 November 2021
doi: 10.3389/fnsys.2021.752219

Direct Approach or Detour: A Comparative Model of Inhibition and Neural Ensemble Size in Behavior Selection

Trond A. Tjøstheim, Birger Johansson and Christian Balkenius*

Department of Philosophy, Lund University Cognitive Science, Lund, Sweden

Organisms must cope with different risk/reward landscapes in their ecological niche. Hence, species have evolved behavior and cognitive processes to optimally balance approach and avoidance. Navigation through space, including taking detours, appears also to be an essential element of consciousness. Such processes allow organisms to negotiate predation risk and natural geometry that obstruct foraging. One aspect of this is the ability to inhibit a direct approach toward a reward. Using an adaptation of the well-known detour paradigm in comparative psychology, but in a virtual world, we simulate how different neural configurations of inhibitive processes can yield behavior that approximates characteristics of different species. Results from simulations may help elucidate how evolutionary adaptation can shape inhibitive processing in particular and behavioral selection in general. More specifically, results indicate that both the level of inhibition that an organism can exert and the size of neural populations dedicated to inhibition contribute to successful detour navigation. According to our results, both factors help to facilitate detour behavior, but the latter (i.e., larger neural populations) appears to specifically reduce behavioral variation.

Keywords: detour task, egocentric navigation, allocentric navigation, navigational strategy selection, consciousness, inhibition

OPEN ACCESS

Edited by:
Jon Mallatt,
Washington State University,
United States

Reviewed by:
Grzegorz Juszczak,
Institute of Genetics and Animal Biotechnology, Polish Academy of Sciences (PAN), Poland
Maria Santacà,
University of Padova, Italy

****Correspondence:***
Trond A. Tjøstheim
trond_arild.tjostheim@lucs.lu.se

Received: *02 August 2021*
Accepted: *29 September 2021*
Published: *09 November 2021*

Citation:
Tjøstheim TA, Johansson B and Balkenius C (2021) Direct Approach or Detour: A Comparative Model of Inhibition and Neural Ensemble Size in Behavior Selection. Front. Syst. Neurosci. 15:752219. doi: 10.3389/fnsys.2021.752219

1. INTRODUCTION

Navigation through space, including taking detours, is an essential element of consciousness (Klein and Barron, 2016; Mallatt et al., 2021). Therefore, exploring the basic mechanisms of these behaviors contributes to the study of consciousness, even if the early steps in the evolution of animal navigation were algorithmic-like and lacking in subjective consciousness like in the model presented in this study. When an organism can no longer follow gradients but must use memory and map-like cognitive structures to cope with an environment, that organism comes closer to supporting a representation of space that is not centered on itself. That is, it supports allocentric representations in addition to self-centered, or egocentric representations. The former affords to see the self in relation to the environment, like being *behind* a tree or *to the east* of a river. The latter affords direct movement like going *forward* or turning to the *right*.

Natural environments may require a diverse number of behavioral strategies to yield optimal access to resources, while balancing safety and competition concerns. However, this variety can often be condensed into two major types mentioned above; allocentric map-based navigation or

egocentric direct approach (Bottini and Doeller, 2020). The extent to which species are biased toward egocentric or allocentric navigation is typically dependent on ecological factors like food availability and the availability of sensory cues (Bruck et al., 2017). Much work has been done to compare species with regards to their ability to control the urge to directly approach salient targets like food, mates, or social groups, and be able to navigate around obstacles *via* detour paths (Kabadayi et al., 2018). In psychology and ethology, this kind of behavior is investigated using detour tasks. The idea of these experimental tasks is that an animal cannot directly approach a target, but must navigate or reach around a barrier first (As shown in **Figure 1**). In the case of navigation tasks, there is usually defined a *barrier zone* immediately in front of the barrier, and the time the animal spends in this zone can be used to operationalize an experimental measure of its inhibitory control, which is the ability to inhibit a futile direct approach and then take a detour.

Kabadayi et al. (2018) review how detour tasks are used in animal cognition. They enumerate the various configurations, measurements, and animal species that have so far been employed in this context. According to them, the behaviors of a wide variety of families of species have been measured, including apes (*homo* and hominoidae), monkeys (cercopithecoidae and platyrrhini), lemuriforms (strepsirrhini), canids (canidae), equids (equidae), birds (aves), reptiles (reptilia), amphibians (amphibia), fish (pisces), molluscs (mollusca), and spiders (salticidae). Detour tasks have also been used to elucidate the characteristics of several cognitive capacities that include inhibitory control, insight learning, memory, motor and cognitive development, functional generalization, and social learning.

As mentioned, Kabadayi et al. (2018) enumerate several configurations of the detour task. One of these is the V-shaped semitransparent configuration. This has been used to test social learning, problem solving, and inhibitory control in several canids such as dingos, dogs, and wolves, as well as mammals like mice, and goats, and reptilians like tortoises. For mice, the configuration is typically adapted to have a circular border and be filled with water, while the goal is a platform that allows subjects to escape from submersion. This is in contrast with e.g., canids, where the goal is a reward like food or social interaction. Subjects can either be placed inside the V barrier and having to move out of it (outward task), or outside it, having to move in (inward task). Refer to **Figure 1** for an example of the inward task which is used in this study. The outward task is usually taken to be the more challenging one as it typically requires subjects to move in the opposite direction to the goal.

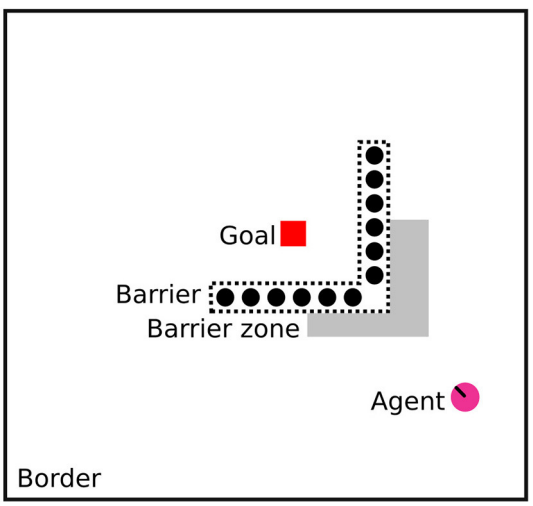


FIGURE 1 | Configuration of the detour task used in experiments, to scale. The barrier is semitransparent with vertical opaque stripes, and the agent is placed facing the apex of the barrier. The diagram shows the goal in red, the semitransparent barrier with transparent parts with dotted lines, the barrier zone in gray, the agent in pink, and the surrounding border.

Focusing on inhibitory ability and behavioral control in the inward, semitransparent V configuration of the detour task, Marshall-Pescini et al. (2015) investigated how wolves (Canis lupus) and dogs (Canis lupus familiaris) differ in this configuration, seeking to test which species can exhibit better inhibitory control. They found that wolves showed shorter latency to reach the goal, and persevered for less time at the barrier. However, dogs had the upper hand in the so-called cylinder task where subjects are required to get at the reward by gaining access through the opening of a cylinder. It is notable that Bray et al. (2015) found that differences appear to exist between dogs with different levels of excitability, or temperament. Comparing calm and excitable dogs, their findings indicate that calm dogs improved their success rate and apparent inhibitory control with increasing arousal, while excitable dogs performed poorer. Juszczak and Miller (2016) employed the V-shaped detour task placed in shallow water to investigate detour behavior in mice. They measured time in the barrier zone in front of the barrier, for both transparent and semitransparent barriers. In their tests, the performance of the mice appeared to depend both on individual inhibitory skills and experience with the task. That is, they found that performance tended to improve over time, and the mice spent less time in the barrier zone as they gained experience.

The ability to change behavior and strategies for approach as presented above is referred to as behavioral flexibility (e.g., Coppens et al., 2010). As the animal studies explain, behavioral flexibility is thought to involve inhibitory activity to balance the influence both of learned behavior and approach motivation toward salient reward stimuli in the immediate environment. For humans, Uddin (2021) identified large-scale functional brain networks encompassing lateral and orbital frontoparietal, midcingulo-insular and frontostriatal regions that support flexibility across the lifespan.

Spiers and Gilbert (2015) propose a conceptual model in which the lateral prefrontal cortex (PFC) provides a prediction error signal about the change in the path, the frontopolar and superior PFC support the re-formulation of the route plan as a novel subgoal and the hippocampus (HC) simulates the new path. Similarly, the ventromedial (vm) PFC may mediate between the conflicting behavioral responses indicated by HC or caudate systems when active (Doeller et al., 2008). The caudate nucleus is involved in landmark-based, egocentric navigation, while HC is

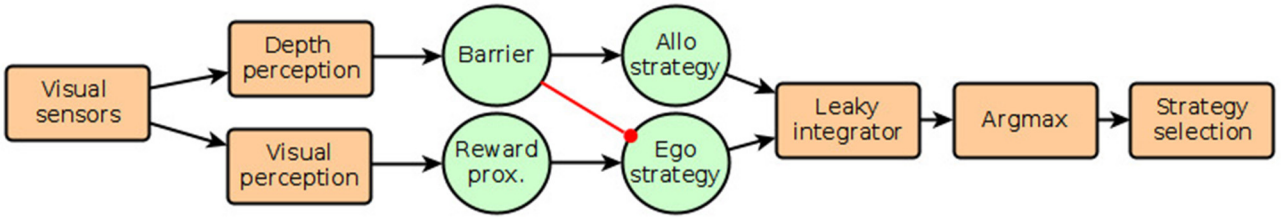


FIGURE 2 | Diagram showing model of strategy selection. Green circular objects represent neural populations that receive signals from perceptual modules. Neural populations are simulated with different numbers of neural units, as described in the text. The red connection indicates inhibition, the level of which is varied between 0 and 1 in simulations. The "Barrier" population is excited by barriers or obstacles immediately in front of the agent, while the "Reward proximity" population is excited by the width of red-colored objects in the visual field.

involved in boundary-based, allocentric navigation. According to Piray et al. (2016), the strength of the vmPFC projection to medial striatum including the caudate nucleus, biases toward model-centric choices. Model-centric strategies are typically associated with allocentric navigation (Doeller et al., 2008). These circuits for navigation present contingent behavioral sequences that can be activated. Which of them will be chosen at any given time is dependent on separate machinery, as explained below.

Neural competition is a cornerstone of many theories of brain function, particularly for processes involved in selection and decision making (Amari, 1977; Grossberg, 1978; Erlhagen and Schöner, 2002). Leaky competing integrator models incorporate aspects of both the psychological and neurophysiological models (Usher and McClelland, 2001, 2004; Johnson and Ratcliff, 2014). Relating to this, Smith (2015) shows that the precision of neural populations increases with the number of participating neural units. In the experiments presented in their study, they used units designed to behave according to an idealized Poisson process, having an exponentially decreasing probability of activity after a stimulus. In the context of visual short term memory, they showed in particular that the signal-to-noise ratio (i.e., the precision) increases proportionally to the square root of the neuronal population size. They also showed that normalization of inputs can be achieved by shunting inhibition, which in practice allows fractional scaling of inputs without losing temporal signatures of signals (Prescott and De Koninck, 2003). According to them, their population-size-dependent normalization model allows theoretical models of reaction time and decision accuracy to be reconciled with experimental data.

Earlier we focused on arousal levels in the context of the noradrenergic system (Balkenius et al., 2018), and found that neural gain in the form of noradrenergic activation may contribute to switching between explorative and exploitative behavioral strategies by e.g., varying the amount of noise present in the selection process. In this study, we concentrate on the effect of varying the size of neural populations, and how that affects precision and integration of sensory information. Additionally, we explore how inhibitive efficacy and precision individually and together can contribute to behavioral strategy selection. Finally, we compare our results with data from experiments on animal species, specifically mice, dogs, and wolves.

2. METHOD

In this section, we explain the rationale behind the model, its properties, and how in particular it is implemented.

2.1. Properties of the Model

To allow selection between the two strategies of egocentric direct approach and allocentric detour, we appropriated a hypothesized network proposed by Barker and Baier (2015). This was originally suggested as a model of approach and avoidance behavior in fish. But given appropriate input signals, it can be used as a winner-takes-all network to select between strategies for approach. In particular, we added one-way inhibition between barrier-collision signals to the neural units representing egocentric strategy. This modified network architecture (as shown in **Figure 2** for a diagram) is informed by work on the spatial pathway from the parietal cortex to vmPFC (Kravitz et al., 2011) that includes boundary sensitive cells in the subiculum (Epstein et al., 2017), and projections from vmPFC to the subthalamic nucleus that can inhibit impulsive behavior (Eagle and Baunez, 2010). The variation of population size and inhibitive strength is likewise informed by Smith (2015) and Piray et al. (2016), respectively.

The spiking model used for the neural elements is as defined by Izhikevich (2003):

$$\frac{dv}{dt} = 0.04v^2 + 5v + 140 - u + I \tag{1}$$

$$\frac{du}{dt} = a(bv - u) \tag{2}$$

$$v = \begin{cases} c, & \text{if } v = 30\text{mV} \\ v, & \text{otherwise} \end{cases} \tag{3}$$

$$u = \begin{cases} u + d, & \text{if } v = 30\text{mV} \\ u, & \text{otherwise} \end{cases} \tag{4}$$

In this study, Equations (1, 2) define pre-spike behavior, while Equations (3, 4) define the reset behavior after a spike. In

TABLE 1 | Numerical values used for simulation.

Parameter	Value
a	0.02
b	0.2
c	$-65.0 + 15\ \gamma^2$
d	$8–6\ \gamma^2$
ω	0.024
ϵ	0.1
λ	0.9
τ	1

Parameters, a, b, c, and d are used for the simulation of spiking units. γ is a noise term between 0 and 1 used to slightly randomize spiking units, as described in Izhikevich (2003). ω, ϵ, λ, and τ are used for the leaky integrator.

Equation (1), I is for direct input current; v is the voltage potential of the unit, and u is a negative feedback variable to v accounting for positive ionic currents. Refer to **Table 1** for parameter values for a, b, c, and d; these values are in accordance with "regular firing" units as defined in Izhikevich (2003).

The formula for the leaky integrator is given by:

$$y_{t+1} = e(x - (1 - l)y_t) \tag{5}$$

where y is the value of the integrator, e is the growth or decay factor (as shown below), x is the input, and l is the leakage factor that affects accumulation. These are defined as follows:

$$e = \begin{cases} \omega, & \text{if } x < \tau \\ \epsilon, & \text{otherwise} \end{cases} \tag{6}$$

$$l = \begin{cases} 0, & \text{if } x < \tau \\ \lambda, & \text{otherwise} \end{cases} \tag{7}$$

Equations (6, 7) define the behavior of the integrator when the input is less than the decay threshold τ. At this point, the integrator begins leaking, or decaying in value, and the value of e changes from ϵ to ω. Refer to **Table 1** for numerical values for these parameters.

2.2. Implementation

The neural simulation model was implemented using the Processing framework v.3.5.3 (Reas and Fry, 2007) with the pOSC library v.0.9.9, while the agent and environment were implemented in the Unity game engine v.3.5. Refer to **Figure 1** for task configuration in Unity. The neural simulation and the agent world were connected using the Open Sound Control (OSC) protocol (Wright and Freed, 1997). In this way, the agent sends out sensory signals while the neural simulation processes these signals, and computes a motor response that is transmitted back and executed by the agent. This back-and-forth communication happens continuously and asynchronously. The set of signals is described in **Table 2**.

The simulation supports two-approach strategies; egocentric direct approach and allocentric approach using a map. The former is implemented by slicing a vector of pixels from the color channels of the cameras, then using pixels from the green and blue channels to remove anything but the purely red pixels in the vector from the red channel. The red pixels are counted, and their center point is calculated. Together, this yields a weighted homing signal that can be used for a direct approach such that the sensor information and the motor signals together form a feedback control circuit.

TABLE 2 | List of OSC messages used to communicate state of agent in simulated environment.

Signal	Description
/camera_r	Red channel from camera
/depth/camera	Depth rendering from camera
/borders	The position and size of the border walls
/obstacles	The position and size of the obstacles
/goals	The position and size of the goal
/agents	The position and size of the agent
/config	An int denoting the current task configuration
/camera/rotation	The relative camera rotation since last step
/camera/absrotation	The absolute camera rotation
/ready	A signal telling the neural simulation that agent is in the initial position and can receive motor commands
/barrierareas	The position and size of the barrier areas

The allocentric map navigation is based on the classical wavefront algorithm (WFA) (Dijkstra, 1959).To facilitate the building of wavefront maps, the agent world sends bounding boxes of all necessary borders, obstacles, and goals, as well as the position of the agent itself. These bounding boxes are used to render a matrix of binary values, making up a map of the environment that can be used by the WFA. The WFA then calculates a gradient from the goal to the agent at every simulation step (to tell if it is getting closer), which gives the agent a direction to move in. This enables the agent to take detours around the obstacles.

As a source of bias for the allocentric strategy, we sliced a vector from the middle of the depth texture from the camera, and transformed it into a two-dimensional matrix. The four topmost rows of this matrix then represent obstacles at various distances from the agent. The rows were weighted and summed up, and the resulting sum was used as a direct input to the spiking population named "Barrier" in **Figure 2**. The spatial pixel density, thus, forms a kind of receptive field similar to those associated with boundary and obstacle cells in medial temporal areas (Epstein et al., 2017; Poulter et al., 2018). Similarly, the aforementioned sum of red pixels taken from the color camera was used as direct input to the parallel spiking population named "Reward proximity" in **Figure 2**. These populations were connected to populations representing either the allocentric strategy or the direct approach strategy, with the output of the obstacle bias also connected to the direct approach unit *via* an adjustable inhibitory weight. Again, refer to **Figure 2** for a diagram of the network. The output of the two strategy units was connected to

TABLE 3 | Summary statistics for simulations with varying population size and inhibition level, listing summary statistics including mean with SD, median with interquartile range (IQR), as well as minimum and maximum values.

Population size	Inhibition	Mean	SD	Median	IQR	Min	Max
1	0.00	26.39	31.03	12.40	8.42	7.40	110.90
1	0.10	11.44	7.82	9.25	5.83	5.50	35.60
1	0.20	15.59	19.95	5.95	4.20	4.40	70.90
1	0.40	10.36	8.34	6.85	5.40	4.30	33.70
1	0.60	5.82	1.49	5.70	1.50	2.20	8.20
1	0.80	5.77	3.88	4.85	0.65	3.20	17.60
1	1.00	5.28	2.19	4.50	1.80	2.10	11.30
2	0.00	15.16	17.91	7.35	4.95	5.80	71.70
2	0.10	6.65	3.21	6.00	0.70	4.80	17.70
2	0.20	11.04	20.32	5.30	0.88	4.60	81.50
2	0.40	7.08	7.46	4.85	1.32	4.10	32.70
2	0.60	10.10	11.62	5.80	3.88	3.70	47.40
2	0.80	6.39	3.64	5.10	1.95	3.70	17.80
2	1.00	5.00	0.99	4.90	1.40	3.60	6.90
5	0.00	13.67	16.34	7.10	2.20	2.40	54.70
5	0.10	5.32	1.02	5.70	1.67	3.30	6.60
5	0.20	5.06	1.83	5.05	0.75	2.00	9.60
5	0.40	4.43	1.46	4.40	0.80	3.00	8.70
5	0.60	9.07	15.84	4.20	1.50	2.10	61.40
5	0.80	7.06	7.94	4.65	1.85	2.10	32.40
5	1.00	5.12	1.76	5.30	1.63	2.00	8.20
10	0.00	9.94	7.17	7.20	3.50	5.30	32.30
10	0.10	5.63	0.64	5.70	1.02	4.70	6.50
10	0.20	5.36	1.32	5.05	0.88	4.00	9.50
10	0.40	5.19	1.11	4.90	0.45	4.00	7.90
10	0.60	4.58	1.32	4.45	1.15	3.00	8.60
10	0.80	4.28	1.12	4.25	1.20	2.80	7.30
10	1.00	3.96	1.12	4.25	1.22	1.20	5.30

leaky integrator units to be able to transform the spiking trains to scalars suitable for identifying the index of the channel with the largest value (argmax selection). This index was then used to select the winning motor commands for transmission to the agent motor system.

During experiments, the level of inhibitory weight was controlled and set to progressively be from zero to one (refer to **Table 3**). The agent was given a starting point in view of the target (refer to **Figure 1**), then left to find its way. The maximum number of steps was set to 1,200, and the simulation was run at 10 Hz, giving a maximum time of 120 s. This makes it possible to compare times in seconds with animal experiments (120 s was also the maximum time limit used for dogs and wolves in Bray et al., 2015). A successful approach to the target was defined as the agent coming within a set radius (5 world units) of the center of the target. After reaching the goal, or the time limit being exceeded, the simulation was reset, parameters for the spiking units were slightly randomized (refer to **Table 1**), and the agent returned to its initial position. Fifteen trials like this were carried out for each inhibitory weight and neuron population size pair. The information gathered from each trial is given again in **Table 3**, and the data was then used to produce statistics.

The statistics was done using Jupyter notebook software (Kluyver et al., 2016), the python programming language (Van Rossum and Drake, 1995), and the Pandas (McKinney, 2010), Seaborn (Waskom, 2021), numpy (Harris et al., 2020), and scipy (Virtanen et al., 2020) libraries.

To calculate the mean and SD of time in the barrier for the animals in **Figure 4**, we used published data from Marshall-Pescini et al. (2015) for dogs and wolves, and Juszczak and Miller (2016) for mice. Our model does not support learning, hence we calculated statistics only for the subset of data that was recorded at the first trial to minimize the effects of learning and experience. Where different barrier configurations were used, we chose only data from the inward-V configuration.

3. RESULTS

In this section, we show results suggesting that increasing the population size of spiking neurons in the neural network generally reduces behavioral variability of the agent, while increasing the weight of inhibition tends to reduce waiting

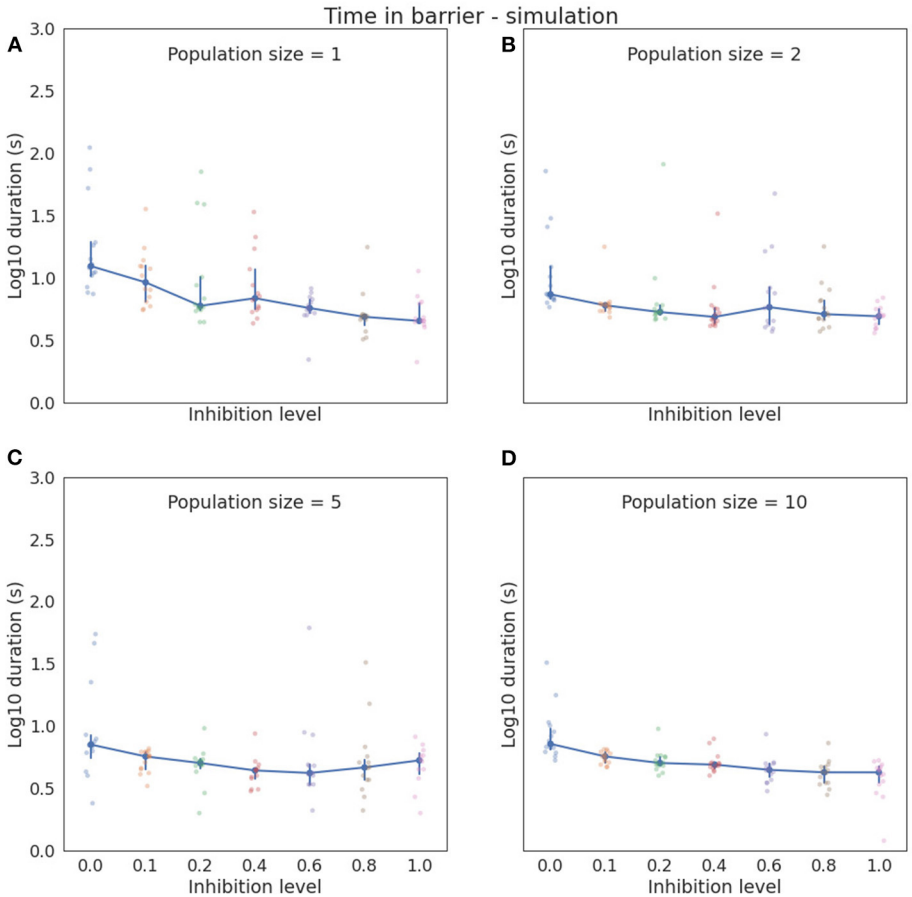


FIGURE 3 | The plot of log10 median time with 95% confidence interval in barrier zone for different simulation configurations. Actual times are indicated by the pale blue and pink dots. **(A)** Simulated neuronal populations each consist of a single neuron. Zero inhibition level yields the highest variance and highest median time in the barrier zone, an while inhibition level of 1 gives the lowest median barrier time. **(B)** Neuronal populations consist of two neurons, **(C)** shows with five neurons, and **(D)** shows with 10 neurons per population. Barrier times and variation generally trend downwards with an increasing number of neurons. Note that median is used instead of mean in these graphs to better accommodate the asymmetric density of the recorded data.

time in the barrier zone. Both of these factors work together to consistently favor the allocentric navigation strategy upon detection of a barrier.

Figure 3 shows barrier wait times for the simulated agents, grouped by inhibition level and the size of the involved neuronal populations. The general trend displayed by this figure is that time in the barrier reduces as the size of the neuronal population grows. Similarly, the variance as indicated by SD reduces. Within a group of the same population sizes, there is an analogous trend of barrier time reduction as inhibition level increases, going from a median of 12.4 (mean = 26.39, SD = 31.03) at zero inhibition and a single neuron per population, to a median of 4.5 (mean = 5.28, SD = 2.19) at inhibition level of one. At the other end of the scale, with 10 neurons per population, the median at zero inhibition is 7.2 (mean = 9.94, SD = 7.17), and 4.25 (mean = 3.96. SD = 1.12) at inhibition of one. It is also noticeable that between the extreme points, both barrier time and variation jump around somewhat for all population sizes except the maximum 10. In this study, the reduction in barrier time is monotonic (as shown in **Table 3**).

Figure 4 shows a scatter plot of mean barrier time vs. SD (i.e., variability). Both animal and simulation data are shown, allowing the animal data to be related to the simulations. Qualitatively, mice spend the least time in the barrier zone and have the least variance, followed by wolves, and with dogs having both the longest time in the barrier, as well as the most variance. Dogs also differ most from the simulated data, spending longer time in the barrier.

4. DISCUSSION

In this final section, we first look at possible explanations for the somewhat surprising position of mouse data on our comparative plot and identify stress as one plausible factor. After that, we turn to the role of inhibition in behavior selection, how the ability to make use of allocentric navigation strategies is an elemental part of consciousness, and how inhibition could be of different use to predator and prey species. We then move to some indications that neural population numbers might not automatically predict

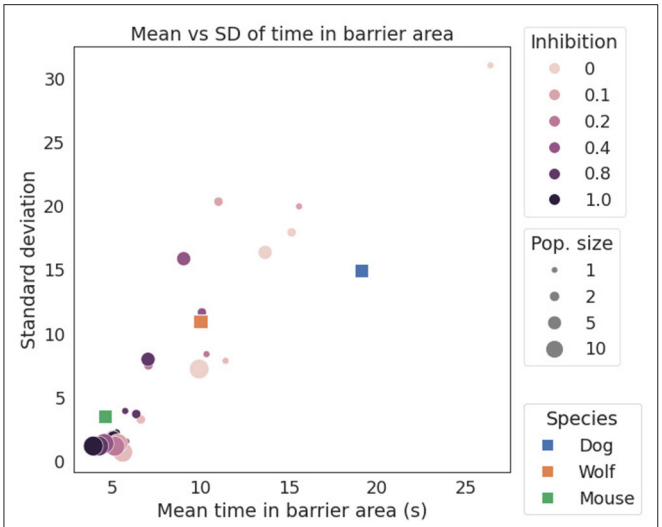


FIGURE 4 | Plot of mean vs. SD for time in barrier zone for different species and simulations. Both the simulation and the animal data appear to be approximately linear. For the three animals, dogs are at the top end of barrier time and variability, and mice at the other extreme. Reducing inhibition yields longer mean time in barrier, while reducing population size increases variability in the form of higher SD. Due to the somewhat stochastic nature of spiking networks, the simulation data naturally displays the noisiness of **Figure 3**.

inhibitive capabilities and discuss how our results might inform findings from animal experiments.

Comparison of behavior between species requires careful controls to take into account differences in anatomy, body structure, and sensory adaptations. Larger bodies tend to require larger brains to control them, and hence direct comparisons of neural numbers are less useful than neural numbers relative to body volume or weight. Another difference between species that can confound comparisons is their dependence on chemical sensation or olfaction. Species for which olfaction is less important are termed *microsmatic*, while those that depend to a large degree on olfaction are termed *macrosmatic* (Santacà et al., 2019a,b). Mice, dogs, and wolves are, hence macrosmatic, while e.g., guppies are considered microsmatic (Santacà et al., 2019a).

One of the interesting inferences one might draw from our results is that mice appear to have more inhibitive powers and larger neuronal populations than do dogs and wolves. One could infer this because mice spend less time at the barrier and more time detouring, so in **Figure 4** they group with the high-inhibition and large-population points. This inference, however, is unlikely to be the actual case. Instead, the reason why mice move out of the barrier zone quickly rather than staying like dogs and wolves could be due to the different experimental designs. Mice are averse to being immersed in water, which is a stressor, and they seek the relief of the above-water platform. This means that the mice engage in escape behavior, or avoidance from an aversive stimulus instead of an approach to a rewarding one, as do dogs and wolves. Furthermore, mice are typically aversive to moving into open spaces, which likely also contributes to them spending less time in the barrier zone (e.g., Bailey and Crawley, 2009). According to Schwabe et al. (2010), mice that were subjected to stress preferred an egocentric strategy more often than an allocentric one. Hence, it would be expected that once a goal is detected, they would engage in a direct approach to that goal and, thus, be likely to persevere at the barrier. But the submerged mice in the detour experiments used the allocentric strategy instead. This demands some further explanation: approach and avoidance activate different behavioral pathways in the brain (Namboodiri et al., 2016), where the avoidance pathways are typically less focused on one particular goal-site and instead result in a kind of "anywhere but here" escape behavior (Gray, 1982; Graeff, 1994). In such panic behavior, animals are even prone to crashing into obstacles in an effort to get away. Gray (1982) argues that the mammalian defense system is hierarchical, with the undirected escape system as the most basic one, and which is active at the most acute level of stress. At lower arousal levels with no stress or panic, the behavioral hierarchy allows goal-directed escape. Some support for this hypothesis might come from Juszczak and Stryjek (2019). They found that administering scopolamine to mice tended to increase perseverance behavior and time in the barrier zone. Given that scopolamine inhibits cholinergic activity by antagonistically binding to muscarinic receptors (Birdsall et al., 1978), and that the cholinergic system contributes to the level of arousal, e.g., in fight or flight behavior (Skinner et al., 2004), one interpretation is that the lowered arousal level induced by scopolamine reduces escape motivation enough that the water-stress configuration used for mice becomes more similar to the approach to reward configuration used for other species; i.e., allowing more decision time at the barrier and more time variance in making the decision to detour. Together these factors might explain the surprising position of mice in **Figure 4**.

Figure 4 shows an approximately linear relationship between mean barrier delay and its variance: more neurons correlate with more inhibition and less delay in successfully choosing to detour. This is in agreement with findings from the animal cognition literature that brain size and neuronal density tend to accompany success rate in tasks that require inhibition (Herculano-Houzel, 2017). Hence, biological neural population numbers can be compared at least relatively to simulated population sizes. This yields the prediction that unstressed mice should display more behavioral variability than dogs in an approach oriented version of the semitransparent V-shaped detour task (i.e., mice in a food-seeking version on dry ground).

Escape behaviors can be automatic, or stimulus-response processes in animals. Such processes are generally believed to be less reliant on consciousness than those necessary for making detours. Consciousness seems to depend on back-and-forth (recurrent) communication between neurons and on the resultant rhythmic synchronization and resonance (e.g., Engel and Singer, 2001; Meador et al., 2002; Engel and Fries, 2016). However, in our model, there are no recurrent connections, and neural populations are not synchronized with rhythmic inhibition. Additionally, as described above, the simulated populations have randomized parameters to explicitly increase activation variance. Hence, there is no direct correlation between neural population activity, and populations are not synchronized.

Therefore, the model indicates that synchronizing populations is not necessary to achieve useful signal integration for behavioral strategy selection in navigation.

Behavioral selection without subjective consciousness also appears to be possible through subcortical pathways to the amygdala. These pathways are held to be evolutionarily older than cortical pathways and are found in both fish and reptiles, as well as mammals (McHaffie et al., 2005). For vision, one such pathway projects from the retina, *via* the brainstem superior colliculus and the thalamic pulvinar nucleus, to the amygdala. This pathway is generally assumed to be responsible for phenomena like blindsight, where people with cortical blindness can still guess the position of objects in their near environment. In particular, signals indicating dangerous stimuli, like the presence of snakes and spiders (and angry faces), are mediated *via* this pathway to the amygdala, which can then engage defensive behaviors. Furthermore, it appears that even routine, non-escape behavior like touching the position of a light signal may be supported by subcortical pathways, without requiring conscious perception. This is evidenced by studies on monkeys (Cowey and Stoerig, 1997).

How could we go from a simple, nonconscious allocentric navigation strategy (**Figure 2**) to one that uses consciousness? Merker (2007) argues that consciousness functionally can be understood *via* a "tripartite" division into (i) target selection (ii) action selection, and (iii) motivational ranking. Although these functions may operate on their own, they typically interact such that motivational ranking can influence target selection, which again can influence the selection of actions. Merker (2007) further argues that these functions need to operate in real time, and that they are integrated *via* a form of simulation. It is this simulation process that effectively constitutes conscious experience. Both target and action selection processes are related to spatial cognition and allow an animal to cope with spatially distributed resources, e.g., that shelter, food, and mates are not all found in the same place. As mentioned above, allocentric maps particularly support navigation to targets that are not directly approachable, or even in the direct vicinity. Hence, a system that allows an animal to be conscious of resource-place associations that are spread out potentially provides evolutionary benefits. Klein and Barron (2016) argue that insect brains may be capable of subjective consciousness since in the proposal of Merker (2007), this is mediated by evolutionary old, subcortical structures like the midbrain and the basal ganglia, and insects have structures that are functionally analogous to these. Similarly, the apparent lack of sufficient spatial perception or sensing in plants is used as a an argument by Mallatt et al. (2021) against plants having consciousness.

Carnivorous predator species and herbivorous prey species have adapted different usage for behavioral inhibition. Whereas, predators could benefit from inhibiting direct approach to prey to avoid detection (Hasson, 1991; Radford et al., 2020), a prey species may use inhibition to stop an approach to potential danger, as well as to "play dead" to reduce attack motivation in a predator (e.g., Gallup et al., 1971). In the case of predators, the perception of an eye pattern in the prey can indicate that the prey is turned in the direction of the predator; this can induce behavioral freeze and change the motivation from a direct approach to detour behavior. This would correspond to the perception of a barrier in our model, and the consequent switch to an allocentric navigation strategy. Similarly, the eyes of predators tend to be front-facing, which is useful for estimating distance (Detwiler, 1955). Prey species, on the other hand, often have side facing eyes since it facilitates surveying larger surrounding areas and hence the detection of potential predators. Although predator and prey species may use inhibition differently to adaptively control behavior, what exactly mediates inhibitory capability in different species is still not completely understood. We turn to this issue next. We have argued above that larger populations of neurons can confer increased precision, but that inhibitive efficacy is not fully dependent on population size. Kabadayi et al. (2017) explored the hypothesis that neuronal population size in the avian pallium might predict success rates on the cylinder task. Given that ravens are very adept at this task, and ravens have a densely populated pallium, they sought to investigate whether other birds with similarly high neural densities perform equally well. Parrots are birds that, like ravens, have comparatively dense palliums. Using parrots as subjects, they did not find evidence for a positive relationship between population size and success on the cylinder task. The parrots performed much poorer than did ravens. The authors interpret these results in two ways. Either that inhibition might not be correlated with pallial neuron count, or that the cylinder task does not measure motor inhibition. Our results lend support to the former of these interpretations (neuron number does not matter in this study) but with a slight twist, namely that there may be differences in inhibitive populations that are independent of total population size but that affect inhibitive efficacy.

Moving from birds to arthropods, Long (2021) compared brain sizes of different spider species and classified the spiders into four groups, where the first group had the smallest brain and the fourth group the largest. Interestingly, a species belonging to the first group, the spitting spider *Scytodes pallidus*, is hunted by a species of the fourth group, the jumping spider *Portia labiata*. Notably, *P. labiata* sometimes changes its hunting strategy depending on whether its prey is a male or female, and whether the female is carrying eggs (Jackson et al., 2002). An egg-carrying female is apparently less dangerous since it must drop its egg to spit. In this case, *P. labiata* makes use of faster, direct-approach strategies. But when hunting a female without eggs, *P. labiata* instead takes longer detours, to attack from behind. This more complex behavior might only be possible due to the larger and more complex brain of *Portia*.

In summary, we have presented a model of navigational strategy selection that shows how a direct approach vs. detour might be influenced by the interplay of both neuronal population size and inhibitive efficacy. The former appears to confer precision that improves signal integration, while the latter facilitates the suppression of direct approach strategies and the usage of allocentric navigation around obstacles. Together both processes contribute to behavioral flexibility in navigating complex environments. Comparing the results presented in this study with data from animal experiences may elucidate differences in inhibitive capabilities in various species.

The work presented in this study opens up several new avenues of exploration and complements earlier simulation work we have presented on awareness (Balkenius et al., 2018) and memory (Balkenius et al., 2020). Combining the present study with the former might further elucidate processes of arousal and how they might affect navigation and behavioral selection in the context of making detours. The latter work on episodic memory and decision making offer exciting opportunities for exploring path-learning and how an agent might react when such paths are changed. In the animal cognition literature, the mechanism by which animals are able to take advantage of shortcuts is an example of this that is of particular interest.

DATA AVAILABILITY STATEMENT

The code for the simulations are publicly available. This data can be found here: https://github.com/trondarild/Tjostheim_et_al_direct_approach_inhibition.

AUTHOR CONTRIBUTIONS

TT conducted simulations and wrote the manuscript in collaboration with and under the supervision of BJ and CB. All authors contributed to the article and approved the submitted version.

FUNDING

This study was partially supported by the Wallenberg AI, Autonomous Systems and Software Program–Humanities and Society (WASP-HS) and funded by the Marianne and Marcus Wallenberg Foundation and the Marcus and Amalia Wallenberg Foundation.

ACKNOWLEDGMENTS

We thank Can Kabadayi for valuable input in planning experiments and writing, as well as the editor and two reviewers for their significant contribution in improving the structure and text of this article.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnsys.2021.752219/full#supplementary-material

REFERENCES

Amari, S.-I. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biol. Cybern.* 27, 77–87. doi: 10.1007/BF00337259

Bailey, K. R., and Crawley, J. N. (2009). "Anxiety-related behaviors in mice," in *Methods of Behavior Analysis in Neuroscience*, 2nd Edn, ed J. J. Buccafusco (Boca Raton, FL: CRC Press).

Balkenius, C., Tjøstheim, T. A., and Johansson, B. (2018). "Arousal and awareness in a humanoid robot," in *CEUR Workshop Proceedings, Vol. 2287 (CEUR Workshop Proceedings)* (Stanford, CA).

Balkenius, C., Tjøstheim, T. A., Johansson, B., Wallin, A., and Gärdenfors, P. (2020). The missing link between memory and reinforcement learning. *Front. Psychol.* 11:3446. doi: 10.3389/fpsyg.2020.560080

Barker, A. J., and Baier, H. (2015). Sensorimotor decision making in the zebrafish tectum. *Curr. Biol.* 25, 2804–2814. doi: 10.1016/j.cub.2015.09.055

Birdsall, N., Burgen, A., and Hulme, E. (1978). The binding of agonists to brain muscarinic receptors. *Mol. Pharmacol.* 14, 723–736.

Bottini, R., and Doeller, C. F. (2020). Knowledge across reference frames: cognitive maps and image spaces. *Trends Cogn. Sci.* 24, 606–619. doi: 10.1016/j.tics.2020.05.008

Bray, E. E., MacLean, E. L., and Hare, B. A. (2015). Increasing arousal enhances inhibitory control in calm but not excitable dogs. *Anim. Cogn.* 18, 1317–1329. doi: 10.1007/s10071-015-0901-1

Bruck, J. N., Allen, N. A., Brass, K. E., Horn, B. A., and Campbell, P. (2017). Species differences in egocentric navigation: the effect of burrowing ecology on a spatial cognitive trait in mice. *Anim. Behav.* 127, 67–73. doi: 10.1016/j.anbehav.2017.02.023

Coppens, C. M., de Boer, S. F., and Koolhaas, J. M. (2010). Coping styles and behavioural flexibility: towards underlying mechanisms. *Philos. Trans. R. Soc. B Biol. Sci.* 365, 4021–4028. doi: 10.1098/rstb.2010.0217

Cowey, A., and Stoerig, P. (1997). Visual detection in monkeys with blindsight. *Neuropsychologia* 35, 929–939. doi: 10.1016/S0028-3932(97)00021-3

Detwiler, S. R. (1955). The eye and its structural adaptations. *Proc. Am. Philos. Soc.* 99, 224–238.

Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numer. Math.* 1, 269–271. doi: 10.1007/BF01386390

Doeller, C. F., King, J. A., and Burgess, N. (2008). Parallel striatal and hippocampal systems for landmarks and boundaries in spatial memory. *Proc. Natl. Acad. Sci. U.S.A.* 105, 5915–5920. doi: 10.1073/pnas.0801489105

Eagle, D. M., and Baunez, C. (2010). Is there an inhibitory-response-control system in the rat? evidence from anatomical and pharmacological studies of behavioral inhibition. *Neurosci. Biobehav. Rev.* 34, 50–72. doi: 10.1016/j.neubiorev.2009.07.003

Engel, A. K., and Fries, P. (2016). "Neuronal oscillations, coherence, and consciousness," in *The Neurology of Consciousness*, eds S. Laureys, O. Gosseries, and G. Tononi (New York, NY: Elsevier), 49–60. doi: 10.1016/B978-0-12-800948-2.00003-0

Engel, A. K., and Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness. *Trends Cogn. Sci.* 5, 16–25. doi: 10.1016/S1364-6613(00)01568-0

Epstein, R. A., Patai, E. Z., Julian, J. B., and Spiers, H. J. (2017). The cognitive map in humans: spatial navigation and beyond. *Nat. Neurosci.* 20, 1504–1513. doi: 10.1038/nn.4656

Erlhagen, W., and Schöner, G. (2002). Dynamic field theory of movement preparation. *Psychol. Rev.* 109, 545. doi: 10.1037/0033-295X.109.3.545

Gallup, G. G., Nash, R. F., and Ellison, A. L. (1971). Tonic immobility as a reaction to predation: artificial eyes as a fear stimulus for chickens. *Psychon. Sci.* 23, 79–80. doi: 10.3758/BF03336016

Graeff, F. G. (1994). Neuroanatomy and neurotransmitter regulation of defensive behaviors and related emotions in mammals. *Braz. J. Med. Biol. Res.* 27, 811–829.

Gray, J. A. (1982). Précis of the neuropsychology of anxiety: an enquiry into the functions of the septo-hippocampal system. *Behav. Brain Sci.* 5, 469–484. doi: 10.1017/S0140525X00013066

Grossberg, S. (1978). Competition, decision, and consensus. *J. Math. Anal. Appl.* 66, 470–493. doi: 10.1016/0022-247X(78)90249-4

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., et al. (2020). Array programming with NumPy. *Nature* 585, 357–362. doi: 10.1038/s41586-020-2649-2

Hasson, O. (1991). Pursuit-deterrent signals: communication between prey and predator. *Trends Ecol. Evolut.* 6, 325–329. doi: 10.1016/0169-5347(91)90040-5

Herculano-Houzel, S. (2017). Numbers of neurons as biological correlates of cognitive capability. *Curr. Opin. Behav. Sci*. 16, 1–7. doi: 10.1016/j.cobeha.2017.02.004

Izhikevich, E. M. (2003). Simple model of spiking neurons. *IEEE Trans. Neural Netw*. 14, 1569–1572. doi: 10.1109/TNN.2003.820440

Jackson, R. R., Pollard, S. D., Li, D., and Fijn, N. (2002). Interpopulation variation in the risk-related decisions of portia labiata, an araneophagic jumping spider (araneae, salticidae), during predatory sequences with spitting spiders. *Anim. Cogn*. 5, 215–223. doi: 10.1007/s10071-002-0150-y

Johnson, E. J., and Ratcliff, R. (2014). "Computational and process models of decision-making in psychology and behavioral economics," in *Neuroeconomics: Decision Making and the Brain*, 2nd Edn, eds P. W. Glimcher and E. Fehr (New York, NY: Academic Press).

Juszczak, G. R., and Miller, M. (2016). Detour behavior of mice trained with transparent, semitransparent and opaque barriers. *PLoS ONE* 11:e0162018. doi: 10.1371/journal.pone.0162018

Juszczak, G. R., and Stryjek, R. (2019). Scopolamine increases perseveration in mice subjected to the detour test. *Behav. Brain Res*. 356, 71–77. doi: 10.1016/j.bbr.2018.07.028

Kabadayi, C., Bobrowicz, K., and Osvath, M. (2018). The detour paradigm in animal cognition. *Anim. Cogn*. 21, 21–35. doi: 10.1007/s10071-017-1152-0

Kabadayi, C., Krasheninnikova, A., O'neill, L., van de Weijer, J., Osvath, M., and von Bayern, A. M. (2017). Are parrots poor at motor self-regulation or is the cylinder task poor at measuring it? *Anim. Cogn*. 20, 1137–1146. doi: 10.1007/s10071-017-1131-5

Klein, C., and Barron, A. B. (2016). Insects have the capacity for subjective experience. *Anim. Sent*. 1, 1. doi: 10.51291/2377-7478.1113

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., et al. (2016). "Jupyter Notebooks-a publishing format for reproducible computational workflows," in *Positioning and Power in Academic Publishing: Players, Agents and Agendas: Proceedings of the 20th International Conference on Electronic Publishing* (Amsterdam: IOS Press), 87.

Kravitz, D. J., Saleem, K. S., Baker, C. I., and Mishkin, M. (2011). A new neural framework for visuospatial processing. *Nat. Rev. Neurosci*. 12, 217–230. doi: 10.1038/nrn3008

Long, S. M. (2021). Variations on a theme: morphological variation in the secondary eye visual pathway across the order of araneae. *J. Compar. Neurol*. 529, 259–280. doi: 10.1002/cne.24945

Mallatt, J., Blatt, M. R., Draguhn, A., Robinson, D. G., and Taiz, L. (2021). Debunking a myth: plant consciousness. *Protoplasma* 258, 459–476. doi: 10.1007/s00709-020-01579-w

Marshall-Pescini, S., Virányi, Z., and Range, F. (2015). The effect of domestication on inhibitory control: wolves and dogs compared. *PLoS ONE* 10:e0118469. doi: 10.1371/journal.pone.0118469

McHaffie, J. G., Stanford, T. R., Stein, B. E., Coizet, V., and Redgrave, P. (2005). Subcortical loops through the basal ganglia. *Trends Neurosci*. 28, 401–407. doi: 10.1016/j.tins.2005.06.006

McKinney, W. (2010). "Data structures for statistical computing in python," in *Proceedings of the 9th Python in Science Conference*, Vol. 445, eds S. van der Walt and J. Millman (Austin, TX: SciPy), 51–56.

Meador, K. J., Ray, P. G., Echauz, J. R., Loring, D. W., and Vachtsevanos, G. J. (2002). Gamma coherence and conscious perception. *Neurology* 59, 847–854. doi: 10.1212/WNL.59.6.847

Merker, B. (2007). Consciousness without a cerebral cortex: a challenge for neuroscience and medicine. *Behav. Brain Sci*. 30, 63–81. doi: 10.1017/S0140525X07000891

Namboodiri, V. M. K., Rodriguez-Romaguera, J., and Stuber, G. D. (2016). The habenula. *Curr. Biol*. 26, R873–R877. doi: 10.1016/j.cub.2016.08.051

Piray, P., Toni, I., and Cools, R. (2016). Human choice strategy varies with anatomical projections from ventromedial prefrontal cortex to medial striatum. *J. Neurosci*. 36, 2857–2867. doi: 10.1523/JNEUROSCI.2033-15.2016

Poulter, S., Hartley, T., and Lever, C. (2018). The neurobiology of mammalian navigation. *Curr. Biol*. 28, R1023–R1042. doi: 10.1016/j.cub.2018.05.050

Prescott, S. A., and De Koninck, Y. (2003). Gain control of firing rate by shunting inhibition: roles of synaptic noise and dendritic saturation. *Proc. Natl. Acad. Sci. U.S.A*. 100, 2076–2081. doi: 10.1073/pnas.0337591100

Radford, C., McNutt, J. W., Rogers, T., Maslen, B., and Jordan, N. (2020). Artificial eyespots on cattle reduce predation by large carnivores. *Commun. Biol*. 3, 1–8. doi: 10.1038/s42003-020-01156-0

Reas, C., and Fry, B. (2007). *Processing: a Programming Handbook for Visual Designers and Artists*. Mit Press.

Santacà, M., Busatta, M., Lucon-Xiccato, T., and Bisazza, A. (2019a). Sensory differences mediate species variation in detour task performance. *Anim. Behav*. 155:153–162. doi: 10.1016/j.anbehav.2019.05.022

Santacà, M., Busatta, M., Savaşçı, B. B., Lucon-Xiccato, T., and Bisazza, A. (2019b). The effect of experience and olfactory cue in an inhibitory control task in guppies, poecilia reticulata. *Anim. Behav*. 151, 1–7. doi: 10.1016/j.anbehav.2019.03.003

Schwabe, L., Schächinger, H., de Kloet, E. R., and Oitzl, M. S. (2010). Corticosteroids operate as a switch between memory systems. *J. Cogn. Neurosci*. 22, 1362–1372. doi: 10.1162/jocn.2009.21278

Skinner, R. D., Homma, Y., and Garcia-Rill, E. (2004). Arousal mechanisms related to posture and locomotion: 2. ascending modulation. *Progr. Brain Res*. 143, 291–298. doi: 10.1016/S0079-6123(03)43028-8

Smith, P. L. (2015). The poisson shot noise model of visual short-term memory and choice response time: normalized coding by neural population size. *J. Math. Psychol*. 66, 41–52. doi: 10.1016/j.jmp.2015.03.007

Spiers, H. J., and Gilbert, S. J. (2015). Solving the detour problem in navigation: a model of prefrontal and hippocampal interactions. *Front. Hum. Neurosci*. 9:125. doi: 10.3389/fnhum.2015.00125

Uddin, L. Q. (2021). Cognitive and behavioural flexibility: neural mechanisms and clinical considerations. *Nat. Rev. Neurosci*. 22, 167–179. doi: 10.1038/s41583-021-00428-w

Usher, M., and McClelland, J. L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychol. Rev*. 108, 550. doi: 10.1037/0033-295X.108.3.550

Usher, M., and McClelland, J. L. (2004). Loss aversion and inhibition in dynamical models of multialternative choice. *Psychol. Rev*. 111, 757. doi: 10.1037/0033-295X.111.3.757

Van Rossum, G., and Drake, F. L. Jr. (1995). *Python Tutorial (Vol. 620)*. Amsterdam: Centrum voor Wiskunde en Informatica.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in python. *Nat. Methods* 17, 261–272. doi: 10.1038/s41592-020-0772-5

Waskom, M. L. (2021). seaborn: statistical data visualization. *J. Open Source Softw*. 6, 3021. doi: 10.21105/joss.03021

Wright, M., and Freed, A. (1997). "Open Sound Control: A New Protocol for Communicating with Sound Synthesizers," in *Proceedings of the 1997 International Computer Music Conference* (San Francisco, CA: International Computer Music Association), 101–104.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Systems Neuroscience

HYPOTHESIS AND THEORY
published: 26 November 2021
doi: 10.3389/fnsys.2021.765646

The Efference Copy Signal as a Key Mechanism for Consciousness

*Giorgio Vallortigara**

Centre for Mind/Brain Sciences, University of Trento, Rovereto, Italy

Animals need to distinguish sensory input caused by their own movement from sensory input which is due to stimuli in the outside world. This can be done by an efference copy mechanism, a carbon copy of the movement-command that is routed to sensory structures. Here I tried to link the mechanism of the efference copy with the idea of the philosopher Thomas Reid that the senses would have a double province, to make us feel, and to make us perceive, and that, as argued by psychologist Nicholas Humphrey, the former would identify with the signals from bodily sense organs with an internalized evaluative response, i.e., with phenomenal consciousness. I discussed a possible departure from the classical implementation of the efference copy mechanism that can effectively provide the senses with such a double province, and possibly allow us some progress in understanding the nature of consciousness.

Keywords: efference copy, corollary discharge, consciousness, sensation/perception, sensory reafference

OPEN ACCESS

Edited by:
Louis Neal Irwin,
The University of Texas at El Paso, United States

Reviewed by:
Xing Tian,
New York University Shanghai, China
Birgitta Dresp-Langley,
Centre National de la Recherche Scientifique (CNRS), France

***Correspondence:**
Giorgio Vallortigara
giorgio.vallortigara@unitn.it

Received: *27 August 2021*
Accepted: *04 November 2021*
Published: *26 November 2021*

Citation:
Vallortigara G (2021) The Efference Copy Signal as a Key Mechanism for Consciousness. Front. Syst. Neurosci. 15:765646. doi: 10.3389/fnsys.2021.765646

INTRODUCTION

> *La música, los estados de felicidad, la mitología, las caras trabajadas por el tiempo, ciertos crepúsculos y ciertos lugares, quieren decirnos algo, o algo dijeron que no hubiéramos debido perder, o están por decir algo; esta inminencia de una revelación, que no se produce, es, quizá, el hecho estético.*
>
> *Jorge Luis Borges*

Since its description by von Holst and Mittelstaedt (1950) and Sperry (1950), the idea that the efference copy signal may play a crucial role in consciousness has been put forward by several authors (see for an historical account Grüsser, 1995; Fukutomi and Carlson, 2020).

The concept of an efference copy arose in the framework of the problem of space constancy, i.e., the fact that the visual world appears stable despite shifts of overall visual input with eye movements. Anticipations of the idea can be found in several authors, such as Bell (1823), Purkinje (1825), von Helmholtz (1866), von Helmholtz and Southall (1962), and von Uexküll (1920), (see Koenderink, 2015) but the breakthrough came from seminal experiments by Erich von Holst and Roger Sperry.

von Holst and Mittelstaedt (1950) inverted the head of the blowfly *Eristalis*, holding it with a piece of wax. The fly appeared to circle either clockwise or counterclockwise at random. Given that in the darkness the fly's movement looked pretty normal, they argued for the existence of a mechanism that compared the output of the locomotor system with the retinal flow field. von Holst and Mittelstaedt (1950) hypothesized an «*Efferenzkopie*» that would be compared and subtracted from the retinal signal to stabilize locomotion. Tilting the head converted the ordinary negative feedback of the efference copy into a positive feedback—a motor command in one direction would feed back a signal to correct in the same direction, thus giving rise to further deviation in the same direction and continuous circling as a result. Sperry (1950) made similar observations in

an independent way, studying fish with surgically inverted eyes, and named the signal «corollary discharge». Although distinctions have been proposed in the literature for use of the two terms (Li et al., 2020), in this article I will use efference copy and corollary discharge interchangeably.

The efference copy signal may enable organisms that move to discount sensory stimulation that arises from their own actions, thereby allowing them to distinguish between the sensory stimulation caused by external stimuli and that caused by their own movements.

Irwin Feinberg (1978) first suggested that failures of the efference copy mechanisms may underlie some of the symptoms of psychosis. This was then developed by Frith (1987) and Shergill et al. (2005). Specifically, Feinberg (1978) argued that dysfunction of efference copy mechanisms that normally allow us to recognize and disregard stimulation resulting from our own actions would characterize schizophrenia, giving rise to the subtle but pervasive sensory/perceptual aberrations observed in these patients. Disturbances of the efference copy mechanisms may contribute to symptoms such as hallucinations and delusions: a failure to recognize one's voice or inner speech as self-generated might produce the subjective experience of an externally generated sound, thus giving rise of auditory hallucination of hearing voices; or a failure to predict the sensory consequences of one's actions may result in the subjective experience of being under the control of external forces.

The mechanisms of the efference copy was then slowly absorbed into the general framework of predictive coding with the idea that the brain needs to infer the causes of a given sensory input, which can be achieved through combining new sensory data with pre-existing knowledge of the world or priors (Ford and Mathalon, 2019). However, several authors have stressed a specific role of efference copy mechanisms on the origins of consciousness (Merker, 2005; Godfrey-Smith, 2016, 2020; Vallortigara, 2021a).

In a recent article, Jékely et al. (2021) argued for a role of *Reafference*, i.e., any effect on an organism's sensory mechanisms that is due to the organism's own actions, to the evolution of the *body-self*, a form of organization that would enable an animal to sense and act as a single unit. The authors noted that reafference in general does not necessarily involve a nervous system: self-initiated activities tend to have predictable consequences, and reafference would simply represent feedbacks concerning such predictions. An example they discussed comes from sponges, in which sensory cilia keep track of the flow produced within the body and can signal when this flow ceases (Ludeman et al., 2014). They argued for a further evolution of the mechanism of reafference when, in animals with nervous systems, sensory and effector devices made available a more sophisticated engine that compensates for predicted sensory changes by registering the particular action underway at a time.

What is unclear in all these accounts is how reafference or efference copy can give rise to consciousness, i.e., to the feelings that accompany and characterize (at times) our responding to sensory stimulation. I believe some progress on this issue can be made if we try to link the idea of the efference copy with the old-fashioned distinction between sensation and perception of some philosophical traditions.

SENSATION AND PERCEPTION

In the *Essays on the Intellectual Powers of Man* Thomas Reid (1941) says that «*When I smell a rose, there is in this operation both sensation and perception. The agreeable odour I feel, considered by itself without relation to any external object, is merely a sensation… Its very essence consists in being felt; and when it is not felt it is not. There is no difference between the sensation and the feeling of it—they are one and the same thing… in sensation there is no object distinct from the act of the mind by which it is felt-and this holds true with regard to all sensations (pp. 150–151)*».

Of course, the terms sensation/perception are associated with a long tradition of debates and different meanings in philosophy (see e.g., Reeves and Dresp-Langley, 2017) but here I am considering only the particular conception developed by this author because of its possible links with biological facts. According to Reid «*The external senses have a double province —to make us feel, and to make us perceive. They furnish us with a variety of sensations, some pleasant, others painful, and others indifferent; at the same time they give us a conception and an invincible belief of the existence of external objects… Sensation, taken by itself, implies neither the conception nor belief of any external object. It supposes a sentient being, and a certain manner in which that being is affected; but it supposes no more. Perception implies a conviction and belief of something external—something different both from the mind that perceives, and the act of perception. Things so different in their nature ought to be distinguished*» (Reid, 1895 [1785], II, Ch. 17 and 16).

Consider the classical example by Reid. When we smell a rose there would be two separate but parallel things happening; namely we feel the sweet smell as a conscious experience (sensation) and we detect the external presence of the object rose (perception). Reid (1895) [1785], II, Ch. 17 and 16) argues that we do not notice or attend to our sensations except under rather special circumstances: «*The mind has acquired a confirmed and inveterate habit of inattention to them, for they no sooner appear than quick as lightning the thing signified succeeds, and engrosses all our regard. They have no name in language; and although we are conscious of them when they pass through the mind, yet their passage is so quick and so familiar, that it is absolutely unheeded* (pp. 135)».

Humphrey (1992, 2006, 2011) beautifully conceptualized the distinction between sensation and perception in terms of representing «what is happening to me» (the feeling of the smell of the rose) and «what is happening out there» (the perception of the object rose). He agrees with Reid that for the most part we overlook our sensations because we focused on the objects of perception. There are, however, clinical conditions that made the sensation/perception distinction apparent. This has been worked out by Humphrey himself, starting from his seminal discovery of the blindsight phenomenon while studying recovering of visual function in the blind monkey Helen (Humphrey and Weiskrantz, 1967). Blindsight patients

can recognize «what is happening out there» but their perception is not accompanied by any conscious feeling, i.e., they lack sensation or the «what is happening to me» (Humphrey, 1992).

Humphrey also moved further from Reid in arguing that having a sensation is not a passive condition but rather a form of active engagement with the stimulus occurring at the body surface. He wrote «*When, for example, I feel pain in my toe, or taste salt on my tongue, or equally when I have red sensation at my eye, I am in effect reaching out to the site of stimulation with a kind of evaluative response—a response appropriate to the stimulus and the body part affected. Indeed what I experience as my sensation of "what is happening to me" is based not on the incoming information as such but rather on the signals I myself am issuing to make the response happen*» (Humphrey, 2000).

THE PRINCIPLE OF REAFFERENCE AS THE FOUNDATION OF THE SENSATION/PERCEPTION DISTINCTION

There are then two questions. First, why should a distinction between sensation and perception be necessary in evolutionary terms? Second, what sort of mechanism can support the distinction between sensation and perception?

As to the first point, the crucial role of active movement has been stressed as lying at the origin of the development of nervous systems (e.g., Llinás, 2001). Active movement also implies the kind of problem that makes necessary the development of an efference copy. As stated by Merker (2005): «*Consider the worm's initiation of a crawling movement. Such a movement will produce sudden stimulation of numerous cutaneous receptors (. . .), yet no withdrawal reflex is released to abort the movement. Apparently the worm's simple nervous system discounts cutaneous stimulation contingent on self-produced movement as a stimulus for withdrawal*».

Thus, one can see the problem of distinguishing «what is happening to me» from «what is happening out there» as a selective pressure that arose specifically with active movement, and the efference copy as the mechanism which has developed through natural selection as a solution of this specific problem.

So far so good but it remains quite a puzzle why sensation (following Reid and Humphrey) should be associated with consciousness (Note that I am referring here to consciousness—which is a word with high polysemy—as simply «experience», i.e., following Block (1995): «*Phenomenal consciousness is experience; the phenomenally conscious aspect of a state is what it is like to be in that state*».). If we take the model of the efference copy we can easily understand why the sensory signal produced by a local stimulation can be annihilated when an efference copy is generated as a result of the active movement of the organism; however, we cannot understand why a sensation would be there in the absence of any active movement, for when an object is impinging on our surface we do feel something (something happening to us).

My proposal is simply to take seriously the hypothesis put forward by Reid and Humphrey and link it with a sort of reversed principle of reafference (see also Hesslow, 2012 for a similar reversed principle, though not linked to sensation and experience). Essentially, the principle of reafference establishes that the organism is able to predict the sensory consequence of its own action, that is, the stimulation that might occur as a result of its own movement. However, one could also consider the situation the other way around: that the body is able to predict the type of motor consequence, that is, of bodily reaction, which should follow from its sensory activity. Indeed, this is exactly what happens, if we assume that the sensation is actually a bodily reaction, a motor action in itself. The double province of the senses might be established by an efference copy of the motoric aspect (the bodily reaction) of the response to the stimulus. Let's examine this hypothesis in more detail in the next section.

CONSCIOUSNESS AS IMMINENCE OF A REVELATION

In the traditional view, the efference copy is a solution to the problem of maintaining the stability of the visual world. So, when for instance an organism moves its eyes, the sliding of retinal images would be canceled by the efference copy associated with the motor command sent to the eye muscles.

Let's consider a slightly different mechanism, arising from some simple experimental phenomenology as shown in **Figures 1A,B**. When we move actively an arm to encounter an object, such as the small pyramid in **Figure 1A**, the active tactile stimulation on the finger is usually associated with the perception of something (an object) out there.

It is quite difficult in these circumstances to focus instead on the feeling of something *on* the finger (which agrees with Reid idea that we do not usually notice or attend to our sensations; and see also more recently Kilteni et al., 2020).

In the reverse condition, however, when the object is moved and hits the finger passively stimulating it, we usually feel something happening to the finger, something happening to us, a sensation (**Figure 1B**).

It seems to me that this can be conceptualized by arguing that sensory stimulation has indeed a double province, namely that the sensory signal is usually associated with a carbon copy of it (an efference copy) which is escorting the sensory signal thus giving rise, as a bodily action, to a sense of agency, i.e., to the fact that such a sensory signal is produced by the organism itself for it is a motor action, a bodily response. If the touching is the result of an active movement of the arm, then the motor signal associated with this movement would nullify the efference copy (the bodily signal) of the local stimulation. The sensory signal would emerge in this case naked from the comparator, giving rise to a perception (something out there) without any sensation (something happening to me). On the contrary, the impinging stimulation caused by the motion of the object itself that hits the finger would be not associated with any cancellation of the efference copy (bodily signal), thus charging the sensory signal of a sense of authorship, what

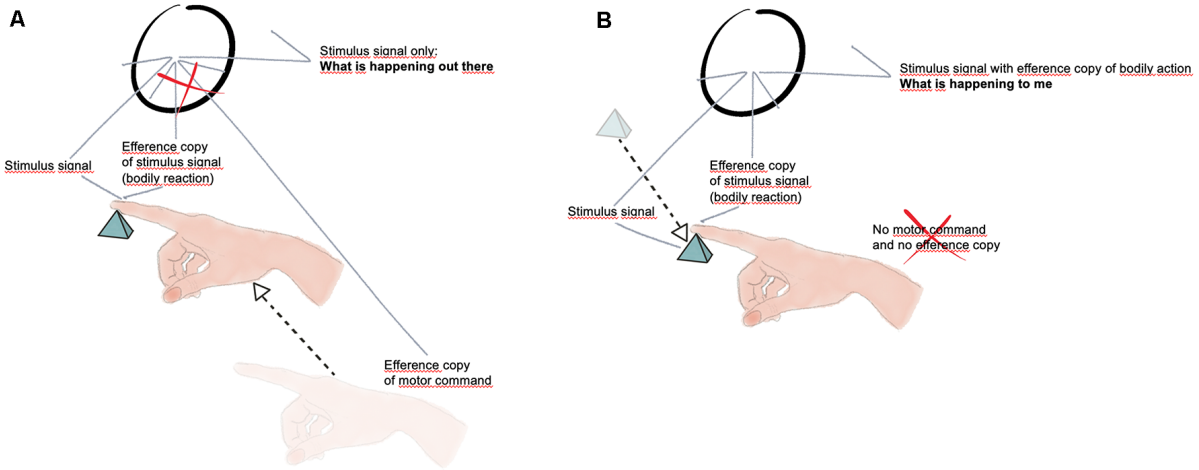


FIGURE 1 | (A) The movement of the hand is associated with an efference copy that annihilates the efference copy associated with the local bodily reaction, thus giving rise to the perception of an object out there without sensation (Drawing by Elena Lorenzi). **(B)** The object is moving and hits the finger; in the absence of active hand movement there is no efference copy to annihilate the efference copy associated with the local bodily reaction, thus a sensation (what is happening to me) arises (Drawing by Elena Lorenzi).

we describe as feeling or experiencing something (The lack of a sense of authorship is probably a crucial aspect of the behavior of blindsight patients, that need to be convinced «to guess»—such a «motivation/reason for action» could have been another basic outcome of the appearance of the double province of senses.).

Although the model would fit with phenomenal experience for tactile stimulation, it may appear a little paradoxical with distant senses: Do we sometimes really not see (in the sense of sensing, feeling it) when looking at the visual world? Well, yes, certainly we do not sense (feel) anything during saccades, i.e., again when the efference copy associated with the bodily action of visual sensing [of «sentition» as Humphrey (1992) dubbed it] is nullified by the efference copy associated with saccadic movements.

Of course, I am not arguing here that the mechanism (nowadays) is peripheral and local. In the scheme argued for by Humphrey (1994), the body's senses produced a local response on the body surface in early organisms but then the response becomes targeted on the incoming sensory nerves and finally privatized in an internal brain circuit. However, my point here is that if the local bodily reaction is not associated with a carbon copy of it to be compared with others motor command as it happens in actively moving organisms, no sensation and no feeling (consciousness) would exist. Similarly, I would not expect sensation to occur in sessile organisms (Vallortigara, 2021b).

Borges wrote (see original text *in esergo*) that «*imminence of a revelation, which is not produced, is, perhaps, the aesthetic event*». This can be used as a metaphor for the reafference theory of consciousness described here, i.e., as a sensory signal which is waiting for a bodily action revelation that may or may not occur (Vallortigara, 2020, 2021a). The operating of the comparator (schematized by the circle in **Figures 1A,B**) that takes into account the different signals likely needs a delay line for the sensory signal of the sort that have been hypothesized in mechanisms such as the Reichardt detector (see Hassenstein and Reichardt, 1956). This time delay could be the foundation of the minimum time duration of the experienced present, an idea dating back to William James (1890) who stressed on the necessity for neural activity to have a suitable duration in order for consciousness to arise from sensory stimulation.

There are advantages in hypothesizing that the comparator would operate on two motoric signals rather than on a sensory and a motoric signal as in the traditional view of the reafference principle (see e.g., for vision Bridgeman, 2010), for we can account better for the phenomenology of our experience and avoid issues that arose with different models of consciousness. Consider for example the ideas put forward by Taylor (1999) who has tried to use the idea of a temporal delay in another way, assuming that the efference copy signal is retained in a temporal memory and that its brief permanence, before its annihilation, would constitute consciousness. In order to do this, Taylor introduced the hypothesis that the corollary discharge is no longer simply derived from the motor signal, but from attention. This corollary discharge of the movement of attention would be retained in a working memory by supplying the properties of experience to the sensory signal before being canceled by it (see Taylor, 2002, 2003).

According to Taylor's model, consciousness is identified with an efference copy of the attention movement control signal residing briefly in its buffer until the associated attended input activation is also arriving in the buffer. The difficulty, however, is that the attributes of the experience in this framework do not seem to belong to the sensory signal itself, but to the corollary discharge (or to the attentional movement control

signal of it). In our example of the hand or the object that moves, the sense of ownership, and of being the agent (the author) of the sensation, would therefore refer to the movement of the finger (or to the attention to the movement of the finger) rather than to the sensation encountered. And in the event that the hand does not move at all but instead is the finger that is passively stimulated by the object due to a displacement and a contact produced by the object itself, there would be no sensation because no attentional movement control signal arises, though sensation is actually happening. Of course, one can argue that besides the efference copy as a potential attentional source, other canonical forms of attention (as heavily investigated in the literature, not necessarily related to motor activity) would be available and thus that the inference from Taylor's theory to no sensation in the absence of no movement would be probably unfair. Nonetheless, claiming for an efference copy of the movement of attention would be problematic also because evidence suggest that consciousness can be observed without attention, and *vice versa* (Koch and Tsuchiya, 2012). These difficulties dissolve, however, if we evaluate the sensory signal for what it is, or better for what it must have been originally as hypothesized by Humphrey (1992), namely a bodily reaction—a movement in itself—with the possibility of making of a carbon copy of it, in the form of an efference copy.

DISCUSSION

In general terms, the reafference principle refers to any kind of effect on an organism's sensory mechanisms that is due to the organism's own actions. It clearly requires some form of motion of the body but as noted by Jékely et al. (2021) «*even a sessile animal can act with reafferent consequences, as when a filter-feeding animal generates a feeding current by motile cilia*». Yet, it seems to me that only the more advanced form of reafference claimed for by Jékely et al. (2021) can be associated with sensation (as opposed to perception), and thus with consciousness. Single cell organisms such as bacteria can use motility to assess the presence of a chemical gradient. Jékely et al. (2021) describe for example a simple form of deformational reafference with an internal reciprocal influence between the sensory events and the effector. However, it is only with the appearance of specialized sensors and effectors that there would be a specific neural signal to convey reafferent sensing during action. In the example I discussed in **Figure 1** involving active touch there is certainly deformational reafference, changes in the shape of the body (at the finger) that lead to sensing. But in order for this sensing to be felt, i.e., to be a sensation, a minimal structure with a sensory neuron, a motor neuron and an interneuron is needed to allow the signal provided by the sensory neuron to be charged (or not to be charged) with the carbon copy (the efference copy) of the motor signal (the deformational bodily reaction) thus providing it with a sense of agency and authorship.

Mechanisms of efference copy have been described at several levels in both vertebrates and invertebrates (Crapse and Sommer, 2008). I would be inclined to consider their presence as a signature of the ability of these organisms to inhabit, as proposed by Reid, a double province of sensory stimulation, that of sensation and that of perception, or in Humphrey's terminology of «what is happening to me» and «what is happening out there». Of course, all this tells us nothing about the specific contents of the sensations of others organisms. Animals with efference copy mechanisms, I would maintain, should be phenomenally conscious, though the contents of their sensations may be incommensurable to each other, for their origins lay in their species-specific bodily reactions on their different body districts.

Objections can be raised of course to the idea that the double province of the senses might be established by an efference copy of the motoric aspect (the bodily reaction) of the response to the stimulus, and several theoretical aspects certainly need more elaboration. Consider the following examples (see e.g., Owen, 2017 for a review on these topics).

First, mental imagery. There is no stimulus during mental imagery. However, according to the cognitive neuroscience literature of mental imagery, the nervous system would be activated similarly as processing a stimulus. How would mental imagery fit in the distinction of «sensation» and «perception», and how does an efference copy contribute to mental imagery? Second, anesthesia would cause dissociation of action and sensation. Would anesthesia produce an illusion of «sensation» and «perception» that are indistinguishable? Third, an extreme case is the locked-in patients who completely lose movement ability. Would the locked-in patients not smell a rose?

I believe that with respect to these three examples we need to consider the changes that occurred in evolutionary history. At the start sensation was a bodily reaction at the very surface of early organisms (with its efference copy), but then, as stressed by Humphrey (1992, 2000) the local response has become privatized, first by targeting it to incoming sensory nerves and then being entirely located into the brain. Consider again in this regard Feinberg's (1978) ideas about psychosis: thought processes themselves can be considered as motor actions, as argued by Hughlings Jackson (1958), because, I would say, they are retaining their characteristics of an, albeit privatized, bodily reaction and thus have an efference copy, the lack of which may produce schizophrenic symptoms (the patient is no longer the author of the bodily reaction, i.e., the author of his own thoughts). Thus, imagery, anesthesia and lock-in do not pose a problem for feeling something, assuming that there is an internal motor command that is the internalized version of the original bodily reaction at the organism's surface.

Several other important issues remain of course unanswered. For example, Reid's definition of perception does involve some difficulties (see Reeves and Dresp-Langley, 2017). How does one know that an animal believes in the object in front of it? It seems unlikely that fixation of belief is exclusively human. Alex, an African Gray Parrot could tell in a sort of vocal labeling resembling English what he experienced and believed to be present, even including perceptual illusions (Pepperberg, 2002). However, a variety of perceptual illusions have been investigated in non-human animals using traditional

motor responses (Vallortigara, 2004, 2006, 2021c; Rosa-Salva et al., 2014), and there seems to be no reason to assume that these motor responses should have a reduced epistemic value with respect to the vocal labeling of Alex (or, for that matter, with respect to human vocal labeling). Clearly, any further discussion should be placed under the light of insight from animal behavior, since the core assumption of this article implies that animals have evolved in strict association with active movement the beginnings of what we call phenomenal experience.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

Bell, C. (1823). On the motions of the eyes, in illustration of the uses of the muscles and nerves of the orbit. *Philos. Trans. R. Soc.* 113, 166–186. doi: 10.1098/rstl.1823.0017

Block, N. (1995). On a confusion about a function of consciousness. *Behav. Brain Sci.* 18, 227–247. doi: 10.1017/S0140525X00038188

Bridgeman, B. (2010). How the brain makes the world appear stable. *Iperception* 1, 69–72. doi: 10.1068/i0387

Crapse, T. B., and Sommer, M. A. (2008). Corollary discharge across the animal kingdom. *Nat. Rev. Neurosci.* 9, 587–600. doi: 10.1038/nrn2457

Feinberg, I. (1978). Efference copy and corollary discharge: implications for thinking and its disorders. *Schizophr. Bull.* 4, 636–640. doi: 10.1093/schbul/4.4.636

Ford, J. M., and Mathalon, D. H. (2019). Efference copy, corollary discharge, predictive coding and psychosis. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* 4, 764–767. doi: 10.1016/j.bpsc.2019.07.005

Frith, C. D. (1987). The positive and negative symptoms of schizophrenia reflect impairments in the perception and initiation of action. *Psychol. Med.* 17, 631–648. doi: 10.1017/s0033291700025873

Fukutomi, M., and Carlson, B. A. (2020). A history of corollary discharge: contributions of mormyrid weakly electric fish. *Front. Integr. Neurosci.* 14:42. doi: 10.3389/fnint.2020.00042

Godfrey-Smith, P. (2016). *Other Minds: The Octopus and the Evolution of Intelligent Life.* London: Harper and Collins.

Godfrey-Smith, P. (2020). *Metazoa: Animal Life and the Birth of the Mind.* London: Harper and Collins.

Grüsser, O.-J. (1995). "On the history of the ideas of efference copy and reafference.," in *Essays in the History of Physiological Sciences: Proceedings of a Symposium Held at the University Louis Pasteur Strasbourg, on March 26-27th, 1993* (London: The Wellcome Institute Series in the History of Medicine. Clio Medica), 33, 35–56.

Hassenstein, V., and Reichardt, W. (1956). System theoretical analysis of time, sequence and sign analysis of the motion perception of the snout-beetle. *Chlorophanus Zeitschrift für Naturforschung* 11, 513–524.

Hesslow, G. (2012). The current status of the simulation theory of cognition. *Brain Res.* 1428, 71–79. doi: 10.1016/j.brainres.2011.06.026

Humphrey, N. (1992). *A History of the Mind.* New York: Chatto & Windus, Simon & Schuster.

Humphrey, N. (1994). The private world of consciousness. *New Scientist* 1907, 23–25.

Humphrey, N. (2000). "The privatization of sensation," in *The Evolution of Cognition*, eds C. Heyes and L. Huber (Cambridge, MA: MIT Press), 241–252.

Humphrey, N. (2011). *Soul Dust: The Magic of Consciousness.* Princeton, NJ: Quercus Publishing, Princeton University Press.

Humphrey, N. (2006). *Seeing Red: A Study in Consciousness.* New York, NY: Belknap Press/Harvard University Press.

Humphrey, N. K., and Weiskrantz, L. (1967). Vision in monkeys after removal of the striate cortex. *Nature* 215, 595–597. doi: 10.1038/215595a0

Jackson, J. H. (1958). *Selected Writings.* (Vol. 1), ed J. Taylor (New York: Basic Books, Inc.), 366–384.

James, J. (1890). *The Principles of Psychology.* New York, NY: Henry Holt and Company.

Jékely, G., Godfrey-Smith, P., and Keijzer, F. (2021). Reafference and the origin of the self in early nervous system evolution. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 376:20190764. doi: 10.1098/rstb.2019.0764

Kilteni, K., Engele, P., and Ehrsson, H. H. (2020). Efference copy is necessary for the attenuation of self-generated touch. *iScience* 23:100843. doi: 10.1016/j.isci.2020.100843

Koch, C., and Tsuchiya, N. (2012). Attention and consciousness: related yet different. *Trends Cogn. Sci.* 16, 103–105. doi: 10.1016/j.tics.2011.11.012

Koenderink, J. (2015). Ontology of the mirror world. *Gestalt Theory* 37, 119–140.

Li, S., Zhu, H., and Tian, X. (2020). Corollary discharge versus efference copy: distinct neural signals in speech preparation differentially modulate auditory responses. *Cereb. Cortex* 30, 5806–5820. doi: 10.1093/cercor/bhaa154

Llinás, R. R. (2001). *I of the Vortex: From Neurons to Self.* Cambridge, MA: MIT Press.

Ludeman, D. A., Farrar, N., Riesgo, A., Paps, J., and Leys, S. P. (2014). Evolutionary origins of sensation in metazoans: functional evidence for a new sensory organ in sponges. *BMC Evol. Biol.* 14:3. doi: 10.1186/1471-2148-14-3

Merker, B. (2005). The liabilities of mobility: a selection pressure for the transition to consciousness in animal evolution. *Conscious. Cogn.* 14, 89–114. doi: 10.1016/S1053-8100(03)00002-3

Owen, A. M. (2017). *Into the Grey Zone: A Neuroscientist Explores the Border Between Life and Death.* New York, NY: Scribner.

Pepperberg, I. M. (2002). Cognitive and communicative abilities of grey parrots. *Curr. Direct. Psychol. Sci.* 11, 83–87. doi: 10.1111/1467-8721.00174

Purkinje, J. (1825). *Beobachtungen und Versuche zur Physiologie der Sinne. Neue Beiträge zur Kenntniss des Sehens in subjectiver Hinsicht.* Berlin: Reimer.

Reid, T. (1941). *Essays on the Intellectual Powers of Man*, ed A. D. Woozley. London: Macmillan and Co.

Reid, T. (1895). *An Inquiry Into the Human Mind on the Principles of Common Sense*, The Philosophical Works of Thomas Reid, 8th edition, ed Sir William Hamilton (Castle Donington, United Kingdom), I, 114.

Reeves, A., and Dresp-Langley, B. (2017). Perceptual categories derived from Reid's "common sense" philosophy. *Front. Psychol.* 8:893. doi: 10.3389/fpsyg.2017.00893

Rosa-Salva, O., Sovrano, V. A., and Vallortigara, G. (2014). What can fish brains tell us about visual perception? *Front. Neural Circuits* 8:119. doi: 10.3389/fncir.2014.00119

Shergill, S. S., Samson, G., Bays, P. M., Frith, C. D., and Wolpert, D. M. (2005). Evidence for sensory prediction deficits in schizophrenia. *Am. J. Psychiatry* 162, 2384–2386. doi: 10.1176/appi.ajp.162.12.2384

Sperry, R. W. (1950). Neural basis of the spontaneous optokinetic response produced by visual inversion. *J. Comp. Physiol. Psychol.* 43, 482–489. doi: 10.1037/h0055479

Taylor, J. G. (1999). *The Race for Consciousness.* Cambridge, MA: MIT Press.

Taylor, J. G. (2003). Paying Attention to Consciousness. *Prog. Neurobiol.* 71, 305–335. doi: 10.1016/j.pneurobio.2003.10.002

Taylor, J. G. (2002). Paying attention to consciousness. *Trends Cogn. Sci.* 5, 206–210. doi: 10.1016/S1364-6613(02)01890-9

Vallortigara, G. (2004). "Visual cognition and representation in birds and primates," in *Vertebrate Comparative Cognition: Are Primates Superior to Non-Primates?*, eds L. J. Rogers and G. Kaplan (New York, NY: Kluwer Academic/Plenum Publishers), 57–94.

Vallortigara, G. (2006). "The cognitive chicken: visual and spatial cognition in a non-mammalian brain," in *Comparative Cognition: Experimental Explorations of Animal Intelligence*, eds E. A. Wasserman and T. R. Zentall (Oxford, UK: Oxford University Press), 41–58.

Vallortigara, G. (2020). Lessons from miniature brains: cognition cheap, memory expensive (sentience linked to active movement?). *Anim. Sentience* 29. doi: 10.51291/2377-7478.1603

Vallortigara, G. (2021a). *Pensieri Della Mosca Con La Testa Storta (Thoughts of the Fly With the Turned Head).* Milan: Adelphi.

Vallortigara, G. (2021b). The rose and the fly. A conjecture on the origin of consciousness. *Biochem. Biophys. Res. Commun.* 564, 170–174. doi: 10.1016/j.bbrc.2020.11.005

Vallortigara, G. (2021c). *Born Knowing.* Cambridge, MA: MIT Press.

von Helmholtz, H. (1866). *Handbuch der physiologischen Optik*. Leipzig: Voss.

von Helmholtz, H., and Southall, J. P. C. (1962). *Helmholtz's Treatise on Physiological Optics*. New York, NY: Dover Publications.

von Holst, E., and Mittelstaedt, H. (1950). The reafference principle: interaction between the central nervous system and the periphery. *Die Naturwissenschaften* 37, 464–476.

von Uexküll, J. (1920). *Theoretische Biologie.* Berlin: Verlag von Gebrüder Paetel.

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

MINI REVIEW
published: 15 December 2021
doi: 10.3389/fpsyg.2021.791398

Current Understanding of the "Insight" Phenomenon Across Disciplines

Antonio J. Osuna-Mascaró and Alice M. I. Auersperg*

Messerli Research Institute, University of Veterinary Medicine, Medical University of Vienna, University of Vienna, Vienna, Austria

Despite countless anecdotes and the historical significance of insight as a problem solving mechanism, its nature has long remained elusive. The conscious experience of insight is notoriously difficult to trace in non-verbal animals. Although studying insight has presented a significant challenge even to neurobiology and psychology, human neuroimaging studies have cleared the theoretical landscape, as they have begun to reveal the underlying mechanisms. The study of insight in non-human animals has, in contrast, remained limited to innovative adjustments to experimental designs within the classical approach of judging cognitive processes in animals, based on task performance. This leaves no apparent possibility of ending debates from different interpretations emerging from conflicting schools of thought. We believe that comparative cognition has thus much to gain by embracing advances from neuroscience and human cognitive psychology. We will review literature on insight (mainly human) and discuss the consequences of these findings to comparative cognition.

Keywords: insight, comparative cognition, problem solving, neuroimaging, comparative psychology

OPEN ACCESS

Edited by:
Lars Chittka,
Queen Mary University of London, United Kingdom

Reviewed by:
Santiago Arango-Munoz,
University of Antioquia, Colombia

****Correspondence:***
Antonio J. Osuna-Mascaró
antonio.osunamascaro@vetmeduni.ac.at

Specialty section:
This article was submitted to Consciousness Research, a section of the journal Frontiers in Psychology

Received: *11 October 2021*
Accepted: *15 November 2021*
Published: *15 December 2021*

Citation:
Osuna-Mascaró AJ and Auersperg AMI (2021) Current Understanding of the "Insight" Phenomenon Across Disciplines. Front. Psychol. 12:791398. doi: 10.3389/fpsyg.2021.791398

INTRODUCTION

A 7 years old girl is standing at a table into which psychologists have fixed a vertical transparent tube containing a small basket with a handle and a sparkly sticker inside. On the table, alongside the tubes, lie a long straight piece of pipe-cleaner and a colorful string. After inserting her finger which only reaches down about a third of the tube, the girl immediately grabs the pipe-cleaner and attempts several times to use it to press the handle of the basket against the tube wall and pull it up. The tube is too narrow and the attempts remain unsuccessful. With a hesitant movement, the colorful string is also briefly dangled into the tube before she seems to get distracted (Isen et al., 1987; Subramaniam et al., 2009). Her gaze seems lost for a moment (Segal, 2004; Kohn and Smith, 2009) when suddenly her pupils dilate (Salvi et al., 2020) and a smile appears (van Steenburgh et al., 2012). She expresses a drawn-out and slightly soaring "Aaahhhh!" and immediately grabs the pipe-cleaner, bends a little hook into one of its distal ends, inserts the hooked end of the pipe-cleaner back into the tube, hooks the handle of the basket, pulls the basket over the rim, and claims her reward with determination (Stuyck et al., 2021).

The hook bending paradigm is a so-called ill-structured innovation task in which the path to the solution is missing information about how to get from its start to its goal state (Cutting

et al., 2014). Interestingly, children that are seven or older find the entire multistep solution to this problem very suddenly rather than in an incremental way. Notably, the hook bending task has similarly been used to test tool innovation in large brained birds and apes, which show a rather ratchet-like improvement upon solving the task for the first time (rarely failing after first success; Weir, 2002; Bird and Emery, 2009a; Laumer et al., 2017, 2018).

The moment just before the little girl tackles the problem, or what Hermann von Helmholtz referred to as a "happy idea" (Wallas, 1926), may be a familiar sentiment to most of us. Such moments of so-called insight are also a recurringly described (and romanticized) phenomenon in scientific history: Newton and that apple, Archimedes in the bathtub, and Poincaré stepping on the bus; all of them have a common pattern: someone with accumulated experience escapes for a moment from the problem to be solved and suddenly finds themselves surprised (without knowing how or why) with the solution.

INSIGHT AS A GLOBAL PHENOMENON

Although there are cultural differences in the importance we attribute to insight as a source of creative output (Rudowicz and Yue, 2000; Niu and Sternberg, 2006; Shao et al., 2019), the traditional description of the stages of the creative process is very similar in European psychology (four stage model by Wallas, 1926) and Eastern philosophy (Yoga Sutras; Maduro, 1976; Shao et al., 2019). Insight itself also has an important bearing in Eastern cultures. For example, in Theravada Buddhism, the goal of vipassana meditation is to reach a sudden understanding, abhisamaya (insight), which contrasts with gradually attained understanding (anapurva). Both the description of the phenomenon and the way in which it is achieved, fit with the popular Western notion of insight (Laukkonen and Slagter, 2021).

Although we can have reasonable confidence that insight is a global phenomenon and not a myth specific to western culture (a WEIRD one; Henrich et al., 2010), it still holds many mysteries regarding its mechanisms and function (Shen et al., 2018), as well as its evolution and presence (and level of expression) in other species (Call, 2013).

SCIENTIFIC INSIGHT

Given the importance of the subjectively perceived components of insight, the phenomenon is certainly easier to study in humans than in non-human animals, both because of the possibility to report verbally (the subject might describe the suddenness of the solution's appearance and the emotions involved, but also specific difficulties with aspects of the task, and how close the subject believes he or she is to the solution at any given moment) and the methodology (because of test diversity and the relative ease of applying neuroimaging technology).

A review by Kounios and Beeman (2014) defines insight as any sudden comprehension, realization, or problem solution that involves a reorganization of the elements of a subject's mental representation of a stimulus, situation, or event to yield a non-obvious or nondominant interpretation. Note, however, that there are various definitions of insight with some considering it as a dynamic process, and others as an end state (Call, 2013; Kounios and Beeman, 2014; Shen et al., 2018). Insight is further frequently linked to a number of traits (such as an impasse or a pleasant feeling of surprise) that may or may not be considered essential to some authors, resulting in variation in the respective definitions (as reviewed in Kounios and Beeman, 2014; and the reason we are using their definition). While neuroscience has been hampered by some inconsistencies in definitions of insight (see Kounios and Beeman, 2014 for examples), experimental evidence (especially due to advances in neuroimaging; e.g., Shen et al., 2018) has helped to guide research along a convergent path (Stuyck et al., 2021), suggesting that innovation achieved through insight-like experiences can be clearly distinguished from other problem solving strategies (van Steenburgh et al., 2012).

Despite the success within neuroscience, the topic of insight and even the use of the term in animal behavior has caused significant theoretical debates in comparative cognition (e.g., Kacelnik, 2009; von Bayern et al., 2009; Emery, 2013). Notably, few animal studies are included the recent literature on human problem solving or neuroscience (Shettleworth, 2012; Call, 2013).

FIRST SCIENTIFIC APPROXIMATIONS TO INSIGHT

In 1925–1926, Wolgang Köhler and Graham Wallas independently published two books that had long lasting effects on the general perception of problem solving: The Mentality of Apes, by Köhler, and The Art of Thoughts, by Wallas.

Wallas, inspired by the ideas of Hermann von Helmholtz and Henri Poincare, proposed four stages of progression for a creative process (Wallas, 1926). Helmholtz, during a banquet held for his 70th birthday in 1891, revealed how he had reached his best ideas; always after first researching a problem in detail, letting it rest, and seeking a pleasant distraction. This way he was often surprised by a solution in the form of a pleasant experience. Wallas named these stages preparation (investigative stage), incubation (temporally discarding the problem from conscious thought), and illumination (the sudden arrival of a new "happy idea"), to which he added a fourth, the verification of the solution. These four stages have been recurrently used as a framework for studying insight in the psychological literature (Luo and Niki, 2003; Jung-Beeman et al., 2004; Sandkühler and Bhattacharya, 2008; Weisberg, 2013). Although Wallas' work covers the creative process in rather broad terms, its relevance to the study of insight is remarkable, due to the close proximity and similarity in conceptualization, measures, and processes (Shen et al., 2017, 2018).

Almost at the same time, Wolfgang Köhler, one of the pioneers of Gestalt psychology, introduced the term insight into comparative psychology (although this way of problem solving was already described before him in non-human animals; Turner, 1909; Köhler, 1925; Weisberg, 2006; Galpayage Dona and Chittka, 2020). Gestalt psychologists proposed that insight depends on different mechanisms to trial and error learning, which, according to Thorndike (1911), was the only way in which animals could solve problems (Köhler, 1925; Koffka, 1935; Duncker, 1945; Wertheimer, 1959). Köhler worked for years at the Casa Amarilla in Tenerife (Canary Islands, Spain) with seven chimpanzees, testing them in experiments where they had to find unusual methods to reach food (see **Figure 1**). In those experiments, Köhler found problem solving strategies that did not seem compatible with classical associative learning routines: After an unsuccessful period of trial and error, in which the chimpanzees used familiar strategies, they stopped trying. Nevertheless, after a while some of them returned with a completely different and, this time, immediately successful strategy. After their first success, the animals could immediately retrieve the correct sequence of steps on the following occasions when they faced the same problem. Köhler, at the time, described these strategies as cognitive trial and error and insight, rather than associative processes.

Other Gestalt psychologists adapted Köhler's problem solving methodology to study insight in humans. Duncker (1945), for example, designed situations in which everyday objects had to be used in unusual ways to solve a task (e.g., the candle problem, see **Figure 1**; Duncker, 1945). Notably, if he asked the subjects to use these objects in their usual way before the test, the success rate was reduced. Duncker and other Gestalt psychologists (e.g., Maier, 1930; Luchins, 1942; Scheerer, 1963) concluded that the repeated application of incorrectly selected knowledge could prevent the deep conceptual understanding necessary to achieve insight. This phenomenon is now known as functional fixedness (Duncker, 1945).

It was, however, the British ornithologist W. H. Thorpe who coined in his book Learning and Instinct in Animals (1956) the most prevalent definition of insight in psychology today; "*the sudden production of a new adaptive response not arrived at by trial behaviour or the solution of a problem by the sudden adaptive reorganization of experience*." We will later explain how an over-emphasis on the absence of trial and

FIGURE 1 | (A) The Crow and the Pitcher, illustrated by Milo Winter (1919; Public Domain). Stones must be dropped into water to have access to the liquid, or to a floating object. **(B)** String-pulling; "Still Life with Fruit and a Goldfinch," Abraham Mignon (1660; Public Domain). Goldfinch's detail, right side. To have access to the hanging object, the string must be pulled first; as seem in Jacobs and Osvath (2015). **(C)** Three-boxes experiment; "Grande on an insecure construction" The Mentality of Apes, Köhler (1925; CC) To get the banana, the chimpanzees must pile the boxes. **(D)** Early representation of the nine-dot problem; Egg of Columbus, Sam Loyds Cyclopedia of Puzzles (1914; Public Domain). Nine dots, arranged in three parallel lines, must be linked with four connected straight lines. **(E)** Candle problem; Duncker (1945; Public Domain) A candle must be attached to the wall; subjects are given a box of tacks, a candle, and matches. Problem on top, solution, below. **(F)** Compound Remote Associates Test test; developed by Mednick and Mednick (1967). Subjects are given the three words on top and have to find one to link with each one of them (as the one in brackets). All Public Domain and Creative Commons (CC) images can be found in Wikimedia Commons.

error learning, and a lack of attention to the "reorganization of experience," may have affected the interpretation of insight in comparative cognition.

OUR CURRENT UNDERSTANDING OF INSIGHT

Insight is often conceptualized as a process in which a subject has a sudden realization of how to solve a novel problem (Schooler et al., 1995; Sheth et al., 2009). Thereby specific elements of a subject's mental representation of various stimuli, situations, or events are reorganized to yield a nonobvious or nondominant interpretation (Kounios and Beeman, 2014). Insight is associated with a number of characteristic phases that set it apart from other mental processes employed in problem solving, such as a distinctive subjective momentary experience of surprise and delight, the "aha" or "eureka" moment (Bowden et al., 2005).

Neuroscience typically contrasts insight with analytical reasoning within problem solving. A directly perceivable difference between the two seems to be a more or less gradual progress toward a solution in analytical thinking (Smith and Kounios, 1996), while individuals are abruptly surprised by the latter during an insightful solution (Metcalfe and Wiebe, 1987). Thus, insight is believed to depend by a large degree (but not completely) on unconscious mental processing, as we will see in the next sections (Sandkühler and Bhattacharya, 2008; Shen et al., 2013, 2018; Weisberg, 2013).

Convergent Insight Process Theories

The main theoretical proposals to explain insight largely differ with regards to the amount of conscious processing they describe involved in an insightful event. For example, approaches, such as the representational change theory (also called the redistribution theory; (Ohlsson, 1992, 2011; Knoblich et al., 1999), advocate a completely unconscious redistribution of information (Knoblich et al., 1999; Ohlsson, 2011), whereas the progress monitoring theory (or criterion for satisfactory progress theory; MacGregor et al., 2001; Chu et al., 2007) proposes insight through a conscious process: searching consciously among a pool of possible solutions during which wrongful presumptions are dropped in favor of a working solution.

In an attempt to find a bridge between the strengths of both previous theories, Weisberg proposed an integrated theory of insight comprising several phases: the individual would first attempt to find a solution by using strategies based on long-term memory; if this fails, the subject would use rules of thumb or more complex heuristics to acquire information about the problem before re-confronting its long-term memory; then, a conscious solution *via* a restructuring of old and new information may thereby be achieved; and if the process reaches an impasse and new information is no longer acquired, an unconscious restructuration of knowledge would take place (Weisberg, 2015). Interestingly, the four stages of Weisberg's (2015) proposal bear some parallels to those suggested by Wallas in the mid twentieth century (Wallas, 1926). "Preparation" would comprise the first three phases of the integrated insight theory, while "incubation" and "illumination" could be interpreted as part of the fourth, where insight is achieved through an unconscious process (see above, section four, to find Wallas' proposal).

Fixation and Impasse

The fixation and impasse (the repetition of incorrect strategies, and the following temporary withdrawal of action), as already described by Duncker (1945), are likely the result of an inappropriate knowledge base (Wiley, 1998) or incomplete heuristics (Knoblich et al., 1999, 2001). Knoblich et al. (1999) found that expertise in algebra can negatively affect insightful arithmetic problem solving. Similarly, great apes have trouble innovating a solution to a problem when the tools or objects at their disposal were previously used in a different way (Hanus et al., 2011; Ebel et al., 2020). Such "functional fixedness" may be one of the factors responsible for the fixation leading to an impasse.

It is important to highlight at this point that there are no insight problems but only insight solutions: any problem solved by insight could also be solved analytically (van Steenburgh et al., 2012), and that an impasse (although common) is not required for insight to occur (MacGregor et al., 2001; Ormerod et al., 2002; Kounios and Beeman, 2014). However, the design of a problem is highly important as it determines the nature of its solution/s. Experimental subjects in classical insight challenges, such as Duncker's candle problem (e.g., Duncker, 1945; Knoblich et al., 2001; Huang, 2017), often encounter an impasse prior to the solution. This is much less common in so-called CRAT-based challenges (a specific type of word puzzle, see **Figure 1**; e.g., Cranford and Moss, 2012; Webb et al., 2019) even if they are also solved by insight. This could be because classical tests often have misleading structures and/or contain elements that may provoke functional fixedness (Duncker, 1945; Hanus et al., 2011; Stuyck et al., 2021). Nevertheless, the scientific approach for detecting an impasse may also be problematic (Stuyck et al., 2021): Studies that found no impasse before insightful solutions mostly relied on verbal reports (e.g., Webb et al., 2019), while when other methods were used an impasse was more likely to be detected (e.g., eye tracking, Huang, 2017; neurophysiological measurements, Shen et al., 2018).

Incubation/Restructuring and Illumination

An impasse is usually followed by an incubation/restructuring stage, which is suspected to constitute the insight's core (Wallas, 1926; Sandkühler and Bhattacharya, 2008; Sio and Ormerod, 2009; Cranford and Moss, 2012; Weisberg, 2013). Although restructuring can of course be done consciously (Weisberg, 2015), it may also happen at a time during which a subject consciously withdraws from the problem at hand (van Steenburgh et al., 2012; Kounios and Beeman, 2014; Shen et al., 2018). We know that insight-like responses improve when participants take a break after reaching an impasse (or when the task is

simply removed from their sight; Kohn and Smith, 2009), regardless of the duration of the break, and particularly when the break is occupied with a different, cognitively demanding task; Segal, 2004).

Human neuroimaging and electrophysiology-based studies suggest a significant function of the prefrontal cortex in the process of overcoming impasse to reach incubation (e.g., Qiu et al., 2010; Zhao et al., 2013; Seyed-Allaei et al., 2017; Shen et al., 2018). The right inferior frontal gyrus plays a role in evaluating possible solutions while the left gyrus seems to control the suppression of inappropriate mental sets or dominantly activated associations (e.g., Jung-Beeman et al., 2004; Shen et al., 2013, 2018; Wu et al., 2015). This corresponds with studies reporting brain asymmetries in insight tests. Studies using insight and priming with word hints (where the left hemisphere typically has an advantage; van Steenburgh et al., 2012), the left visual field (right hemisphere) has shown a strong advantage over the right, with primed participants finding more solutions faster (Bowden and Beeman, 1998; Beeman and Bowden, 2000).

Studies based on event-related potentials have so far been able to identify two distinct cognitive processes involved in achieving an insightful event: the breaking down of the impasse (allowing incubation/restructuring) and the formation of new associations prior to the solution (Luo and Niki, 2003; Luo et al., 2011; Zhao et al., 2013; Shen et al., 2018; it is also described as the enlightenment stage by Wallas, 1926).

Associations that will result in a solution can take different routes; once strong yet incorrect associations can be overcome, weaker yet correct association can be detected (Shen et al., 2018). Interestingly, the latter is facilitated by a positive emotional state (Isen et al., 1987; Subramaniam et al., 2009; van Steenburgh et al., 2012). In humans, a positive emotional state at the start of testing is associated with increased activity in the anterior cingulate cortex (which is related to monitoring cognitive conflict; Carter et al., 2000) and an increase in insightful solutions (Subramaniam et al., 2009).

While neurobiology and cognitive psychology embrace insightful solutions achieved by associations learned in the past, comparative cognition tends to exclude associative learning from its notion of insight, which is a misconception as insight can occur through distant or weak associations (Shettleworth, 2012; Call, 2013). In comparative cognition, insight has occasionally been used as a default explanation upon failing to detect the typical gradual process of associative learning.

A candidate for explaining how we can learn non-obvious associations is latent learning (Tolman and Honzik, 1930; Tolman, 1948). The nervous system can register associations without the need for positive reinforcement (such as those that can be acquired through random exploration). These associations remain latent and are candidates for insightful solutions (Thorpe, 1956). Latent associations, being weak, can be adjusted more flexibly if required (Call, 2013). In contrast, strong associations can result in functional fixedness where a previous solution prevents the innovation of a new solution (e.g., humans, Duncker, 1945; great apes, Ebel et al., 2020).

However, the path toward a solution can be achieved by other mechanisms. The free energy principle [the basis of Predictive Processing Theory (PPT), e.g., Hohwy and Seth, 2020; Francken et al., 2021] predicts that all sentient beings minimize uncertainty for energetic reasons (Friston, 2003). According to PPT, all interaction with the environment involves constant amendment between perceptual input and the internal models (Friston et al., 2016a). When the flow of input stops during an impasse, models continue to be optimized without the agent consciously perceiving it. This has been called fact-free learning or model selection and reduction (model selection, Aragones et al., 2005; model reduction, Friston et al., 2016b). In the absence of new data, the only way we can optimize our generative models is by making them simpler (Friston et al., 2017).

Model reduction is a similar process to that described in the N-REM phase of sleep, where redundant connections between neurons are eliminated (Tononi and Cirelli, 2006) and models are reduced in complexity in the absence of new sensory input (Friston et al., 2017).

Model reduction occurs neither only during sleep, nor only in humans. Rats that move away from exploratory or spatial foraging behavior, and enter short periods of rest, have been found to have hippocampal activity similar to what we would expect in models undergoing insight-compatible changes (Gupta et al., 2010; Pezzulo et al., 2014; Friston et al., 2017). Internally generated sequences (sequences of multi-neuron firing activity that do not reflect an ongoing behavioral sequence) seem to be able to restructure models, not only consolidating memory but also exploring potential solutions (Pezzulo et al., 2014).

The Eureka Experience

A popular event related to insight is the so-called "aha" moment, a subjective experience of surprise and delight accompanied by sudden solutions (Bowden et al., 2005; Sandkühler and Bhattacharya, 2008; Weisberg, 2013; Shen et al., 2017). This pleasant experience is probably one of the reasons why insight responses are associated with positive emotions versus analytical solutions that are negatively perceived (Shen et al., 2016, 2017; Webb et al., 2016, 2019). This may also contribute to a better memorization and a higher success rate of insightful responses (e.g., Danek et al., 2013; Webb et al., 2016; Salvi et al., 2020; Stuyck et al., 2021).

Notably, insight does not necessarily require this "aha" experience. In verbal tests, insight lacking major emotional changes has been reported (Kounios and Beeman, 2014). This may be the reason why CRAT tests do not elicit a perceivable impasse experience (Stuyck et al., 2021). Nevertheless, the impasse may be an important contributing factor to the surprise element of the insight revelation as it fosters the perception of a metacognitive error in which we solve a problem faster than expected (Dubey et al., 2021).

The subpersonal nature of model reduction (that is, there is no explicit inner model, hence no conscious experience of the reduction process) could explain why the agent becomes aware at the precise instance of a new association, and not before (Metcalfe and Wiebe, 1987; Friston et al., 2017; Shen et al., 2018). Another proposed explanation for the relation of insight with consciousness is the asymmetrical involvement

of both hemispheres and the important role of the right hemisphere in key parts of the process (see split brain perception studies, e.g., Gazzaniga, 1998; van Steenburgh et al., 2012). Furthermore, the conscious perception of the solution is plausible considering the close relationship between associative learning and consciousness (Ginsburg and Jablonka, 2007, 2019) and the essential role of consciousness for the former to occur (e.g., Baars et al., 2013; Meuwese et al., 2013; Weidemann et al., 2016).

NON-HUMAN ANIMALS, PROBLEMS, AND SOLUTIONS

Comparative cognition has attempted to tackle the presence of insight in animals by rating the speed of their performance on technical problem or their ability to transfer information from one task to another (Seed and Boogert, 2013).

One issue with this may be that, as mentioned earlier, there are no insight problems, only insight solutions; a problem designed to be solved by insight can also be solved by other processes (van Steenburgh et al., 2012). Epstein et al. (1984) tried to highlight this issue in a popular paper which showed that pigeons solved seemingly complex problems spontaneously by "chaining" blocks of previously learned information.

Neuroscience's results and advances have been able to compensate a lack of theoretical consistency regarding insight. Cognitive research on animal insight, on the other hand, has been limited to the creativity of experimental designs, with no apparent chance of ending long-running debates stemming from two opposing schools of thought, cognitive psychology and behaviorism, "romantics" against "killjoys" (Shettleworth, 2010, 2012; Call, 2013; Starzak and Gray, 2021). While we believe that the progress of comparative cognition feeds (as a dissipative structure) on the continued conflict between the two positions, the lack of experimental progress has kept these discussions in an impasse (e.g., Heinrich, 1995; Kacelnik, 2009; Chittka et al., 2012; Taylor et al., 2012; Emery, 2013; Starzak and Gray, 2021).

Today we know that insight is a measurable phenomenon with a physiological basis that is beginning to be revealed (Shen et al., 2018). Moreover, it makes little sense to set the phenomenon apart from associative learning and experience (Shettleworth, 2010, 2012; Hanus et al., 2011; Call, 2013; Shen et al., 2018; Ebel et al., 2020). Insight does not mean developing *de novo* behaviors to solve a problem, but to find a solution by restructuring the problem, even if the agent reorganizes old experiences to apply them to a novel context.

Although insight involves making the nonobvious seem obvious, and even tends to correlate with a higher success rate at problem solving (higher successful rate, Salvi et al., 2016; Webb et al., 2016; but see, Stuyck et al., 2021), a successful restructuring does not necessarily imply a correct conceptualization of the full nature of the problem, and an answer obtained by insight need not necessarily be correct (Kounios and Beeman, 2014). Just as a feeling of understanding does not equate to a true understanding of the problem, we must thus be careful in equating insight with understanding or suggesting that one predicts the other.

Insight may exist in animals outside humans and could even be relatively widespread in nature (e.g., Shettleworth, 2012; Pezzulo et al., 2014). Yet to proficiently tackle the phenomenon in non-verbal species is an unsolved problem in comparative cognition.

While rodent studies suggest that insight does not require sophisticated cognition, the role of the prefrontal cortex in important insight stages may suggest insightful solutions are more likely to emerge in species that have highly developed and functionally equivalent brain regions (Shettleworth, 2010, 2012; Call, 2013; Olkowicz et al., 2016; Shen et al., 2018).

Methodologies, such as the priming of different brain hemispheres, related to insight (which function similarly in non-human primates as in humans) as well as new technologies in animal eye tracking open the door to technically challenging targeted studies in species other than our own (Krupenye et al., 2016; Shen et al., 2018; Völter et al., 2020; Ben-Haim et al., 2021).

The crucial role of subjective experience in insight, as well as the traditional reliance on verbal reports in a large number of studies, makes it tempting to conclude that the study of insight is inaccessible in non-human animals. Nonetheless, other signatures of insight do exist (e.g., Kounios and Beeman, 2014). Apart from EEG and fMRI studies, evidence of human insight stems also from eye tracking studies (e.g., Salvi, 2013; Salvi et al., 2016; Huang, 2017), grip strength (Laukkonen et al., 2021), heart rate (Hill and Kemp, 2018), pupil dilation, and eye movement (with pupil dilation happening only just prior to an insightful event, and an increase in microsaccade rate coinciding with analytic responses; Salvi et al., 2020). Moreover, it has been shown repeatably that agents do not even necessarily need to solve the problem. A promising approach could be to confront an animal with a problem and then, after a period unsuccessful interaction, to suddenly show the solution and record the response (e.g., Kizilirmak et al., 2016; Webb et al., 2019).

Even the "aha" moment itself might be accessible to study in non-verbal subjects, given the expected physiological emotional response that follows it. We know that many animals show an emotional response while learning how to solve tasks (independent from the presence of a reward; e.g., cows, Hagen and Broom, 2004; goats, Langbein et al., 2004; horses, Mengoli et al., 2014; dogs, McGowan et al., 2014; dolphins, Clark et al., 2013). Studying insight through the presentation of a solution would thus require both a behavioral analysis (as in traditional contrafreeloading tests or yoked experimental designs; e.g., Hagen and Broom, 2004; Rosenberger et al., 2020) as well as a physiological one. Artificially altering the transparency of the path toward the solution, and altering the time spent at an apparent impasse, may allow us to predict and modify the intensity of the respective physiological (as it would be an increased heart rate; Hill and Kemp, 2018) and behavioral responses (e.g., in dogs, we would predict pupil dilation, tail wagging, and increased general activity; McGowan et al., 2014; Webb et al., 2019; Salvi et al., 2020).

CONCLUSION

Insight is a measurable phenomenon in humans, and the mechanisms by which it occurs may well be accessible to species other than our own. Thanks to recent progress in neuroscience and human psychology, we are beginning to clarify the (in some cases subtle) differences that distinguish insight problem solving from other processes. Comparative cognition, however, has so far been limited in its approach. Performance-based setups using technical problems in both birds and mammals have produced highly interesting and suggestive, yet, ambivalent evidence on animal insight (e.g., Heinrich, 1995; Mendes et al., 2007; Bird and Emery, 2009a,b; Laumer et al., 2017, 2018; von Bayern et al., 2018). We are optimistic that accomplishments in neuroscience and human psychology over the past decade can be incorporated into and inspire future comparative cognition studies in their ongoing quest to learn about the capacity for insight in species other than our own.

AUTHOR CONTRIBUTIONS

AO-M wrote the first draft. AO-M and AA finished the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

The authors are funded by the WWTF Project CS18-023 and START project Y 01309 by the Austrian Science Fund (FWF) to AA.

ACKNOWLEDGMENTS

We thank Poppy J. Lambert for her helpful suggestions and language correction of the manuscript.

REFERENCES

Aragones, E., Gilboa, I., Postlewaite, A., and Schmeidler, D. (2005). Fact-free learning. *Am. Econ. Rev.* 95, 1355–1368. doi: 10.1257/000282805775014308

Baars, B. J., Franklin, S., and Ramsoy, T. Z. (2013). Global workspace dynamics: cortical "binding and propagation" enables conscious contents. *Front. Psychol.* 4:200. doi: 10.3389/fpsyg.2013.00200

Beeman, M. J., and Bowden, E. M. (2000). The right hemisphere maintains solution-related activation for yet-to-be-solved problems. *Mem. Cogn.* 28, 1231–1241. doi: 10.3758/BF03211823

Ben-Haim, M. S., Dal Monte, O., Fagan, N. A., Dunham, Y., Hassin, R. R., Chang, S. W. C., et al. (2021). Disentangling perceptual awareness from nonconscious processing in rhesus monkeys (*Macaca mulatta*). *Proc. Natl. Acad. Sci.* 118:e2017543118. doi: 10.1073/pnas.2017543118

Bird, C. D., and Emery, N. J. (2009a). Insightful problem solving and creative tool modification by captive nontool-using rooks. *Proc. Natl. Acad. Sci.* 106, 10370–10375. doi: 10.1073/pnas.0901008106

Bird, C. D., and Emery, N. J. (2009b). Rooks use stones to raise the water level to reach a floating worm. *Curr. Biol.* 19, 1410–1414. doi: 10.1016/j.cub.2009.07.033

Bowden, E. M., and Beeman, M. J. (1998). Getting the right idea: semantic activation in the right hemisphere may help solve insight problems. *Psychol. Sci.* 9, 435–440. doi: 10.1111/1467-9280.00082

Bowden, E. M., Jung-Beeman, M., Fleck, J., and Kounios, J. (2005). New approaches to demystifying insight. *Trends Cogn. Sci.* 9, 322–328. doi: 10.1016/j.tics.2005.05.012

Call, J. (2013). "Three ingredients for becoming a creative tool user," in *Tool Use in Animals: Cognition and Ecology*. eds. C. Boesch, C. M. Sanz and J. Call (Cambridge: Cambridge University Press), 3–20.

Carter, C. S., Macdonald, A. M., Botvinick, M., Ross, L. L., Stenger, V. A., Noll, D., et al. (2000). Parsing executive processes: strategic vs. evaluative functions of the anterior cingulate cortex. *Proc. Natl. Acad. Sci. U. S. A.* 97, 1944–1948. doi: 10.1073/pnas.97.4.1944

Chittka, L., Rossiter, S. J., Skorupski, P., and Fernando, C. (2012). What is comparable in comparative cognition? *Philo. Trans. Royal Soc. Biol. Sci.* 367, 2677–2685. doi: 10.1098/rstb.2012.0215

Chu, Y., Dewald, A., and Chronicle, E. (2007). Theory driven hints in the cheap necklace problem: A preliminary investigation. *J. Probl. Solving* 1:4. doi: 10.7771/1932-6246.1010

Clark, F. E., Davies, S. L., Madigan, A. W., Warner, A. J., and Kuczaj, S. A. II. (2013). Cognitive enrichment for bottlenose dolphins (*Tursiops truncatus*): evaluation of a novel underwater maze device. *Zoo Biol.* 32, 608–619. doi: 10.1002/zoo.21096

Cranford, E., and Moss, J. (2012). Is insight always the same? A protocol analysis of insight in compound remote associate problems. *J. Probl. Solving* 4:8. doi: 10.7771/1932-6246.1129

Cutting, N., Apperly, I. A., Chappell, J., and Beck, S. R. (2014). The puzzling difficulty of tool innovation: why can't children piece their knowledge together? *J. Exp. Child Psychol.* 125, 110–117. doi: 10.1016/j.jecp.2013.11.010

Danek, A. H., Fraps, T., von Müller, A., Grothe, B., and Öllinger, M. (2013). Aha! Experiences leave a mark: facilitated recall of insight solutions. *Psychol. Res.* 77, 659–669. doi: 10.1007/s00426-012-0454-8

Dubey, R., Ho, M., Mehta, H., and Griffiths, T. (2021). Aha! Moments correspond to meta-cognitive prediction errors. *PsyArXiv*. doi: 10.31234/osf.Io/c5v42, [Epub Ahead of Print]

Duncker, K. (1945). On problem-solving. *Psychol. Monogr.* 58, 1–113. doi: 10.1037/h0093599

Ebel, S., Völter, C., and Call, J. (2020). Prior experience mediates the usage of food items as tools in great apes (pan paniscus, pan troglodytes, Gorilla gorilla, and Pongo abelii). *J. Comp. Psychol.* 135, 64–73. doi: 10.1037/com0000236

Emery, N. J. (2013). "Insight, imagination and invention: tool understanding in a non-tool-using corvid," in *Tool Use in Animals: Cognition and Ecology*. eds. C. Boesch, C. M. Sanz and J. Call (Cambridge: Cambridge University Press), 67–88.

Epstein, R., Kirshnit, C. E., Lanza, R. P., and Rubin, L. C. (1984). 'Insight' in the pigeon: antecedents and determinants of an intelligent performance. *Nature* 308, 61–62. doi: 10.1038/308061a0

Francken, J., Beerendonk, L., Molenaar, D., Fahrenfort, J., Kiverstein, J., Seth, A., et al. (2021). An academic survey on theoretical foundations, common assumptions and the current state of the field of consciousness science. PsyArXiv. doi: 10.31234/osf.Io/8mbsk, [Epub Ahead of Print].

Friston, K. (2003). Learning and inference in the brain. *Neural. Netw.* 16, 1325–1352.

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., O'Doherty, J., and Pezzulo, G. (2016a). Active inference and learning. *Neurosci. Biobehav. Rev.* 68, 862–879. doi: 10.1016/j.neubiorev.2016.06.022

Friston, K. J., Lin, M., Frith, C. D., Pezzulo, G., Hobson, J. A., and Ondobaka, S. (2017). Active inference, curiosity and insight. *Neural Comput.* 29, 2633–2683. doi: 10.1162/neco_a_00999

Friston, K. J., Litvak, V., Oswal, A., Razi, A., Stephan, K. E., van Wijk, B. C. M., et al. (2016b). Bayesian model reduction and empirical Bayes for group (DCM) studies. *NeuroImage* 128, 413–431. doi: 10.1016/j.neuroimage.2015.11.015

Galpayage Dona, H. S. G., and Chittka, L. (2020). Charles H. Turner, pioneer in animal cognition. *Science* 370, 530–531. doi: 10.1126/science.abd8754

Gazzaniga, M. S. (1998). The Split brain revisited. *Sci. Am.* 279, 50–55. doi: 10.1038/scientificamerican0798-50

Ginsburg, S., and Jablonka, E. (2007). The transition to experiencing: II. The evolution of associative learning based on feelings. *Biol. Theory* 2, 231–243. doi: 10.1162/biot.2007.2.3.231

Ginsburg, S., and Jablonka, E. (2019). *The Evolution of the Sensitive Soul: Learning and the Origins of Consciousness*. Cambridge, MA, United States: The MIT Press.

Gupta, A. S., van der Meer, M. A. A., Touretzky, D. S., and Redish, A. D. (2010). Hippocampal replay is not a simple function of experience. *Neuron* 65, 695–705. doi: 10.1016/j.neuron.2010.01.034

Hagen, K., and Broom, D. (2004). Emotional reactions to learning in cattle. *Appl. Anim. Behav. Sci.* 85, 203–213. doi: 10.1016/j.applanim.2003.11.007

Hanus, D., Mendes, N., Tennie, C., and Call, J. (2011). Comparing the performances of apes (Gorilla gorilla, pan troglodytes, Pongo pygmaeus) and human children (Homo sapiens) in the floating Peanut task. *PLoS One* 6:e19555. doi: 10.1371/journal.pone.0019555

Heinrich, B. (1995). An experimental investigation of insight in common ravens (Corvus corax). *Auk* 112, 994–1003. doi: 10.2307/4089030

Henrich, J., Heine, S. J., and Norenzayan, A. (2010). The weirdest people in the world? *Behav. Brain Sci.* 33, 61–83. doi: 10.1017/S0140525X0999152X

Hill, G., and Kemp, S. M. (2018). Connect 4: A novel paradigm to elicit positive and negative insight and search problem solving. *Front. Psychol.* 9:1755.

Hohwy, J., and Seth, A. (2020). Predictive processing as a systematic basis for identifying the neural correlates of consciousness. *Philo. Mind Sci.* 1:2. doi: 10.33735/phimisci.2020.ii.64

Huang, P.-S. (2017). An exploratory study on remote associates problem solving: evidence of eye movement indicators. *Think. Skills Creat.* 24, 63–72. doi: 10.1016/j.tsc.2017.02.004

Isen, A., Daubman, K., and Nowicki, G. P. (1987). Positive affect facilitates creative problem solving. *J. Pers. Soc. Psychol.* 52, 1122–1131. doi: 10.1037/0022-3514.52.6.1122

Jacobs, I. F., and Osvath, M. (2015). The string-pulling paradigm in comparative psychology. *J. Comp. Psychol.* 129, 89–120. doi: 10.1037/a0038746

Jung-Beeman, M., Bowden, E. M., Haberman, J., Frymiare, J. L., Arambel-Liu, S., Greenblatt, R., et al. (2004). Neural activity when people solve verbal problems with insight. *PLoS Biol.* 2:e97. doi: 10.1371/journal.pbio.0020097

Kacelnik, A. (2009). Tools for thought or thoughts for tools? *Proc. Natl. Acad. Sci.* 106, 10071–10072. doi: 10.1073/pnas.0904735106

Kizilirmak, J., Wiegmann, B., and Richardson-Klavehn, A. (2016). Problem solving as an encoding task: A special case of the generation effect. *J. Probl. Solving* 9:5. doi: 10.7771/1932-6246.1182

Knoblich, G., Ohlsson, S., Haider, H., and Rhenius, D. (1999). Constraint relaxation and chunk decomposition in insight problem solving. *J. Exp. Psychol. Learn. Mem. Cogn.* 25, 1534–1555. doi: 10.1037/0278-7393.25.6.1534

Knoblich, G., Ohlsson, S., and Raney, G. E. (2001). An eye movement study of insight problem solving. *Mem. Cogn.* 29, 1000–1009. doi: 10.3758/BF03195762

Koffka, K. (1935). Principles of Gestalt Psychology. Available at: http://archive.org/details/in.ernet.dli.2015.7888 (Accessed October 3, 2021).

Köhler, W. (1925). *The Mentality of Apes*. Oxford, England: Harcourt, Brace.

Kohn, N., and Smith, S. M. (2009). Partly versus completely Out of your mind: effects of incubation and distraction on resolving fixation. *J. Creat. Behav.* 43, 102–118. doi: 10.1002/j.2162-6057.2009.tb01309.x

Kounios, J., and Beeman, M. (2014). The cognitive neuroscience of insight. *Annu. Rev. Psychol.* 65, 71–93. doi: 10.1146/annurev-psych-010213-115154

Krupenye, C., Kano, F., Hirata, S., Call, J., and Tomasello, M. (2016). Great apes anticipate that other individuals will act according to false beliefs. *Science* 354, 110–114. doi: 10.1126/science.aaf8110

Langbein, J., Nürnberg, G., and Manteuffel, G. (2004). Visual discrimination learning in dwarf goats and associated changes in heart rate and heart rate variability. *Physiol. Behav.* 82, 601–609. doi: 10.1016/j.physbeh.2004.05.007

Laukkonen, R. E., Ingledew, D. J., Grimmer, H. J., Schooler, J. W., and Tangen, J. M. (2021). Getting a grip on insight: real-time and embodied aha experiences predict correct solutions. *Cognit. Emot.* 35, 918–935. doi: 10.1080/02699931.2021.1908230

Laukkonen, R. E., and Slagter, H. A. (2021). From many to (n)one: meditation and the plasticity of the predictive mind. *Neurosci. Biobehav. Rev.* 128, 199–217. doi: 10.1016/j.neubiorev.2021.06.021

Laumer, I. B., Bugnyar, T., Reber, S. A., and Auersperg, A. M. I. (2017). Can hook-bending be let off the hook? Bending/unbending of pliant tools by cockatoos. *Proc. R. Soc. B* 284:20171026. doi: 10.1098/rspb.2017.1026

Laumer, I. B., Call, J., Bugnyar, T., and Auersperg, A. M. I. (2018). Spontaneous innovation of hook-bending and unbending in orangutans (Pongo abelii). *Sci. Rep.* 8:16518. doi: 10.1038/s41598-018-34607-0

Luchins, A. S. (1942). Mechanization in problem solving: The effect of Einstellung. *Psychol. Monogr.* 54, 1–95. doi: 10.1037/h0093502

Luo, J., Li, W., Fink, A., Jia, L., Xiao, X., Qiu, J., et al. (2011). The time course of breaking mental sets and forming novel associations in insight-like problem solving: an ERP investigation. *Exp. Brain Res.* 212, 583–591. doi: 10.1007/s00221-011-2761-5

Luo, J., and Niki, K. (2003). Function of hippocampus in "insight" of problem solving. *Hippocampus* 13, 316–323. doi: 10.1002/hipo.10069

MacGregor, J. N., Ormerod, T. C., and Chronicle, E. P. (2001). Information processing and insight: A process model of performance on the nine-dot and related problems. *J. Exp. Psychol. Learn. Mem. Cogn.* 27, 176–201. doi: 10.1037/0278-7393.27.1.176

Maduro, R. (1976). Artistic creativity in a Brahmin painter community. Undefined. Available at: https://www.semanticscholar.org/paper/Artistic-creativity-in-a-Brahmin-painter-community-Maduro/98d9f9965dc305a5b2a970d355b0f98386c40a84 (Accessed October 3, 2021).

Maier, N. R. F. (1930). Reasoning in humans. I. On direction. *J. Comp. Psychol.* 10, 115–143. doi: 10.1037/h0073232

McGowan, R. T. S., Rehn, T., Norling, Y., and Keeling, L. J. (2014). Positive affect and learning: exploring the "Eureka effect" in dogs. *Anim. Cogn.* 17, 577–587. doi: 10.1007/s10071-013-0688-x

Mednick, S. A., and Mednick, M. T. S. (1967). *Remote Associates Test, College, Adult, Form 1 and Examiner's Manual, Remote Associates Test, College and Adult Forms 1 and 2*. United States: Houghton Mifflin Company.

Mendes, N., Hanus, D., and Call, J. (2007). Raising the level: orangutans use water as a tool. *Biol. Lett.* 3, 453–455. doi: 10.1098/rsbl.2007.0198

Mengoli, M., Pageat, P., Lafont-Lecuelle, C., Monneret, P., Giacalone, A., Sighieri, C., et al. (2014). Influence of emotional balance during a learning and recall test in horses (Equus caballus). *Behav. Process.* 106, 141–150. doi: 10.1016/j.beproc.2014.05.004

Metcalfe, J., and Wiebe, D. (1987). Intuition in insight and noninsight problem solving. *Mem. Cogn.* 15, 238–246. doi: 10.3758/BF03197722

Meuwese, J., Scholte, S., and Lamme, V. (2013). Does perceptual learning require consciousness or attention? *J. Vis.* 13:912. doi: 10.1167/13.9.912

Niu, W., and Sternberg, R. J. (2006). The philosophical roots of Western and eastern conceptions of creativity. *J. Theor. Philos. Psychol.* 26, 18–38. doi: 10.1037/h0091265

Ohlsson, S. (1992). Information-processing explanations of insight and related phenomena. *Adv. Psychol. Think. Chap.* 1, 1–44.

Ohlsson, S. (2011). *Deep Learning: How the Mind Overrides Experience*. New York, NY, United States: Cambridge University Press.

Olkowicz, S., Kocourek, M., Lučan, R. K., Porteš, M., Fitch, W. T., Herculano-Houzel, S., et al. (2016). Birds have primate-like numbers of neurons in the forebrain. *Proc. Natl. Acad. Sci. U. S. A.* 113, 7255–7260. doi: 10.1073/pnas.1517131113

Ormerod, T. C., MacGregor, J. N., and Chronicle, E. P. (2002). Dynamics and constraints in insight problem solving. *J. Exp. Psychol. Learn. Mem. Cogn.* 28, 791–799. doi: 10.1037/0278-7393.28.4.791

Pezzulo, G., van der Meer, M. A. A., Lansink, C. S., and Pennartz, C. M. A. (2014). Internally generated sequences in learning and executing goal-directed behavior. *Trends Cogn. Sci.* 18, 647–657. doi: 10.1016/j.tics.2014.06.011

Qiu, J., Li, H., Jou, J., Liu, J., Luo, Y., Feng, T., et al. (2010). Neural correlates of the "aha" experiences: evidence from an fMRI study of insight problem solving. *Cortex* 46, 397–403. doi: 10.1016/j.cortex.2009.06.006

Rosenberger, K., Simmler, M., Nawroth, C., Langbein, J., and Keil, N. (2020). Goats work for food in a contrafreeloading task. *Sci. Rep.* 10:22336. doi: 10.1038/s41598-020-78931-w

Rudowicz, E., and Yue, X.-D. (2000). Concepts of creativity: similarities and differences among mainland, Hong Kong and Taiwanese Chinese. *J. Creat. Behav.* 34, 175–192. doi: 10.1002/j.2162-6057.2000.tb01210.x

Salvi, C. (2013). Look outside the box, to think outside the box: insight, eye movements and solution accuracy. Milano: Milano-Bicocca University.

Salvi, C., Bricolo, E., Kounios, J., Bowden, E., and Beeman, M. (2016). Insight solutions are correct more often than analytic solutions. *Think. Reason.* 22, 443–460. doi: 10.1080/13546783.2016.1141798

Salvi, C., Simoncini, C., Grafman, J., and Beeman, M. (2020). Oculometric signature of switch into awareness? Pupil size predicts sudden insight whereas microsaccades predict problem-solving via analysis. *NeuroImage* 217:116933. doi: 10.1016/j.neuroimage.2020.116933

Sandkühler, S., and Bhattacharya, J. (2008). Deconstructing insight: EEG correlates of insightful problem solving. *PLoS One* 3:e1459. doi: 10.1371/journal.pone.0001459

Scheerer, M. (1963). Problem-solving. *Sci. Am.* 208, 118–132. doi: 10.1038/scientificamerican0463-118

Schooler, J. W., Fallshore, M., and Fiore, S. M. (1995). "Epilogue: putting insight into perspective," in *The Nature of Insight*, eds. R. J. Sternberg and J. E. Davidson (Cambridge, CA: The MIT Press), 559–588.

Seed, A. M., and Boogert, N. J. (2013). Animal cognition: An end to insight? *Curr. Biol.* 23, R67–R69. doi: 10.1016/j.cub.2012.11.043

Segal, E. (2004). Incubation in insight problem solving. *Creat. Res. J.* 16, 141–148. doi: 10.1207/s15326934crj1601_13

Seyed-Allaei, S., Avanaki, Z. N., Bahrami, B., and Shallice, T. (2017). Major thought restructuring: The roles of different prefrontal cortical regions. *J. Cogn. Neurosci.* 29, 1147–1161. doi: 10.1162/jocn_a_01109

Shao, Y., Zhang, C., Zhou, J., Gu, T., and Yuan, Y. (2019). How does culture shape creativity? *Mini-Rev. Front. Psychol.* 10:1219. doi: 10.3389/fpsyg.2019.01219

Shen, W., Luo, J., Liu, C., and Yuan, Y. (2013). New advances in the neural correlates of insight: A decade in review of the insightful brain. *Chin. Sci. Bullet.* 58, 1497–1511. doi: 10.1007/S11434-012-5565-5

Shen, W., Tong, Y., Li, F., Yuan, Y., Hommel, B., Liu, C., et al. (2018). Tracking the neurodynamics of insight: A meta-analysis of neuroimaging studies. *Biol. Psychol.* 138, 189–198. doi: 10.1016/j.biopsycho.2018.08.018

Shen, W., Yuan, Y., Liu, C., and Luo, J. (2016). In search of the 'aha!' Experience: elucidating the emotionality of insight problem-solving. *Br. J. Psychol.* 107, 281–298. doi: 10.1111/bjop.12142

Shen, W., Yuan, Y., Liu, C., and Luo, J. (2017). The roles of the temporal lobe in creative insight: an integrated review. *Think. Reason.* 23, 321–375. doi: 10.1080/13546783.2017.1308885

Sheth, B. R., Sandkühler, S., and Bhattacharya, J. (2009). Posterior Beta and anterior gamma oscillations predict cognitive insight. *J. Cogn. Neurosci.* 21, 1269–1279. doi: 10.1162/jocn.2009.21069

Shettleworth, S. J. (2010). Clever animals and killjoy explanations in comparative psychology. *Trends Cogn. Sci.* 14, 477–481. doi: 10.1016/j.tics.2010.07.002

Shettleworth, S. (2012). Do animals have insight, and what is insight anyway? *Canadian J. Exp.* 66, 217–226. doi: 10.1037/a0030674

Sio, U. N., and Ormerod, T. C. (2009). Does incubation enhance problem solving? A meta-analytic review. *Psychol. Bull.* 135, 94–120. doi: 10.1037/a0014212

Smith, R. W., and Kounios, J. (1996). Sudden insight: all-or-none processing revealed by speed-accuracy decomposition. *J. Exp. Psychol. Learn. Mem. Cogn.* 22, 1443–1462. doi: 10.1037//0278-7393.22.6.1443

Starzak, T. B., and Gray, R. D. (2021). Towards ending the animal cognition war: a three-dimensional model of causal cognition. *Biol. Philos.* 36, 1–24. doi: 10.1007/s10539-021-09779-1

Stuyck, H., Aben, B., Cleeremans, A., and Van den Bussche, E. (2021). The aha! Moment: is insight a different form of problem solving? *Conscious. Cogn.* 90:103055. doi: 10.1016/j.concog.2020.103055

Subramaniam, K., Kounios, J., Parrish, T. B., and Jung-Beeman, M. (2009). A brain mechanism for facilitation of insight by positive affect. *J. Cogn. Neurosci.* 21, 415–432. doi: 10.1162/jocn.2009.21057

Taylor, A. H., Knaebe, B., and Gray, R. D. (2012). An end to insight? New Caledonian crows can spontaneously solve problems without planning their actions. *Proc. R. Soc. B Biol. Sci.* 279, 4977–4981. doi: 10.1098/rspb.2012.1998

Thorndike, E. L. (1911). *Animal Intelligence: Experimental Studies*. Lewiston, NY, United States: Macmillan Press.

Thorpe, W. H. (1956). *Learning and Instinct in Animals*. Cambridge, MA, United States: Harvard University Press.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychol. Rev.* 55, 189–208. doi: 10.1037/h0061626

Tolman, E. C., and Honzik, C. H. (1930). Introduction and removal of reward, and maze performance in rats. *Univ. Pub. Psychol.* 4, 257–275.

Tononi, G., and Cirelli, C. (2006). Sleep function and synaptic homeostasis. *Sleep Med. Rev.* 10, 49–62. doi: 10.1016/j.smrv.2005.05.002

Turner, C. H. (1909). The behavior of a Snake. *Science* 30, 563–564. doi: 10.1126/science.30.773.563

van Steenburgh, J. J., Fleck, J. I., Beeman, M., and Kounios, J. (2012). *Insight. The Oxford Handbook of Thinking and Reasoning*. United Kingdom: Oxford University Press.

Völter, C. J., Karl, S., and Huber, L. (2020). Dogs accurately track a moving object on a screen and anticipate its destination. *Sci. Rep.* 10:19832. doi: 10.1038/s41598-020-72506-5

von Bayern, A. M. P., Danel, S., Auersperg, A. M. I., Mioduszewska, B., and Kacelnik, A. (2018). Compound tool construction by new Caledonian crows. *Sci. Rep.* 8:15676. doi: 10.1038/s41598-018-33458-z

von Bayern, A. M. P., Heathcote, R. J. P., Rutz, C., and Kacelnik, A. (2009). The role of experience in problem solving and innovative tool use in crows. *Curr. Biol.* 19, 1965–1968. doi: 10.1016/j.cub.2009.10.037

Wallas, G. (1926). *The Art of Thought*. New York: Harcourt, Brace and Company.

Webb, M. E., Cropper, S. J., and Little, D. R. (2019). "Aha!" is stronger when preceded by a "huh?": presentation of a solution affects ratings of aha experience conditional on accuracy. *Think. Reason.* 25, 324–364. doi: 10.1080/13546783.2018.1523807

Webb, M. E., Little, D. R., and Cropper, S. J. (2016). Insight is not in the problem: investigating insight in problem solving across task types. *Front. Psychol.* 7:1424. doi: 10.3389/fpsyg.2016.01424

Weidemann, G., Satkunarajah, M., and Lovibond, P. F. (2016). I think, therefore Eyeblink: The importance of contingency awareness in conditioning. *Psychol. Sci.* 27, 467–475. doi: 10.1177/0956797615625973

Weir, A. A. S. (2002). Shaping of hooks in new Caledonian crows. *Science* 297:981. doi: 10.1126/science.1073433

Weisberg, R. W. (2006). *Creativity: Understanding Innovation in Problem Solving, Science, Invention, and the Arts*. Hoboken, NJ, United States: John Wiley and Sons Inc.

Weisberg, R. W. (2013). On the "demystification" of insight: A critique of neuroimaging studies of insight. *Creat. Res. J.* 25, 1–14. doi: 10.1080/10400419.2013.752178

Weisberg, R. W. (2015). Toward an integrated theory of insight in problem solving. *Think. Reason.* 21, 5–39. doi: 10.1080/13546783.2014.886625

Wertheimer, M. (1959). Productive thinking. New York: Harper Available at: http://books.google.com/books?id=c1N9AAAAMAAJ (Accessed October 3, 2021).

Wiley, J. (1998). Expertise as mental set: The effects of domain knowledge in creative problem solving. *Mem. Cogn.* 26, 716–730. doi: 10.3758/BF03211392

Wu, X., Yang, W., Tong, D., Sun, J., Chen, Q., Wei, D., et al. (2015). A meta-analysis of neuroimaging studies on divergent thinking using activation likelihood estimation. *Hum. Brain Mapp.* 36, 2703–2718. doi: 10.1002/hbm.22801

Zhao, Q., Zhou, Z., Xu, H., Chen, S., Xu, F., Fan, W., et al. (2013). Dynamic neural network of insight: A functional magnetic resonance imaging study on solving Chinese 'Chengyu' riddles. *PLoS One* 8:e59351. doi: 10.1371/journal.pone.0059351

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Systems Neuroscience

HYPOTHESIS AND THEORY
published: 14 March 2022
doi: 10.3389/fnsys.2022.840022

Where Is It Like to Be an Octopus?

*Sidney Carls-Diamante**

Zukunftskolleg/Philosophy Department, University of Konstanz, Konstanz, Germany

The cognitive capacities and behavioural repertoire of octopuses have led to speculation that these animals may possess consciousness. However, the nervous system of octopuses is radically different from those typically associated with conscious experience: rather than being centralised and profoundly integrated, the octopus nervous system is distributed into components with considerable functional autonomy from each other. Of particular note is the arm nervous system: when severed, octopus arms still exhibit behaviours that are nearly identical to those exhibited when the animal is intact. Given these factors, there is reason to speculate that if octopuses do possess consciousness, it may be of a form highly dissimilar to familiar models. In particular, it may be that the octopus arm is capable of supporting an idiosyncratic field of consciousness. As such, in addition to the likelihood that there is something it is like to be an octopus, there may also be something it is like to be an octopus *arm*. This manuscript explores this possibility.

Keywords: octopus consciousness, octopus arm, octo-munculus, multiple consciousness, unity of consciousness

OPEN ACCESS

Edited by:
Nicky S. Clayton,
University of Cambridge,
United Kingdom

Reviewed by:
Andrew M. Haun,
University of Wisconsin-Madison,
United States
Isabel Maria Martin Monzon,
University of Seville, Spain

***Correspondence:**
Sidney Carls-Diamante
sidney.carls-diamante@uni-konstanz.de

Received: *20 December 2021*
Accepted: *17 January 2022*
Published: *14 March 2022*

Citation:
Carls-Diamante S (2022) Where Is It Like to Be an Octopus? Front. Syst. Neurosci. 16:840022. doi: 10.3389/fnsys.2022.840022

INTRODUCTION

It has recently been suggested that the octopus possesses "two brains" (Grasso, 2014). In particular, these are the central brain and the *brachial plexus,* or the network formed by the interconnection of *axial nerve cords*, of which every arm has one. As will be discussed in detail shortly, the axial nerve cords are considered high-level neural centres within each arm, due to their processing and control responsibilities (Richter et al., 2015). The complexity of the octopus's arm nervous system—which makes up the bulk of the peripheral nervous system (PNS)—is such that each arm demonstrates organisation "like the brain of a living organism. . .with a diversity of sensory modalities, motor neurons effecting different motor systems and large central neuropils which are processing centres for large amounts of information" (Grasso, 2014, p. 103). Such features are what prompted suggestions that octopus arms may house local "brains."

Although the brain and arm nervous system are dissimilar in their functions and structure, both make extensive and non-redundant contributions to cognition and behaviour in octopuses. In order to describe the complex interplay between the central and peripheral components of the octopus neuro-cognitive system, Grasso (2014) uses the metaphor of an "octo-munculus" as an illustration. This octo-munculus would be "a brain-to-body spatial map. . .(like the human 'Homunculus'). . .depicted as information processing systems distributed throughout each arm and a brachial centre in the brain" (Shigeno et al., 2018, p. 11).

Now, since philosophy has a long history of associating brains—or in this case sophisticated neural structures with considerable functional autonomy and anatomical demarcation—with minds or consciousness, it is not unreasonable to wonder about what kind of subjective,

phenomenal experience would arise from such a nervous system as that of octopuses. Indeed, the recent years have seen an increase in interest in consciousness in octopuses. Since these animals exhibit behaviour deemed to be indicative of consciousness, yet have nervous systems that are greatly dissimilar from those associated with the capacity to support consciousness, octopuses proffer the possibility of a radically different form of consciousness from what we are currently familiar with. In particular, due to their highly distributed neurocognitive systems with highly autonomous components, the possibility has been raised that octopus consciousness might consist of multiple conscious fields that may or may not be experienced as a single, unified field. This question remains open, and for now I must postpone an attempt at addressing it.

Nevertheless, the very features of the octopus nervous system that suggested disunified consciousness present another potential way wherein octopus consciousness differs from familiar models: octopuses may house two different types of consciousness, with dissimilar complexity, contents, functions, and contributions to cognition. Thus, in addition to speculating about "What it is like to be an octopus?" (see Nagel, 1974), one might also ask "*Where* is it like to be an octopus?". This manuscript explores this latter question, by *raising the possibility* that octopus arms have their own respective conscious fields. In order to achieve this aim, I present a number of principled reasons to surmise about the presence of "arm-based" consciousness in octopuses.

The manuscript proceeds as follows. Section "Consciousness" discusses the construals of consciousness utilised for present purposes. Section "Octopus Nervous System" provides a description of the octopus nervous system and features that are of particular interest. Section "Attributing Consciousness to Octopuses" surveys the bases for consciousness attribution in octopuses. Section "Arm-Based Consciousness" presents the principled reasons for speculating about arm-based conscious fields. Finally, section "Concluding Remarks" concludes the manuscript.

CONSCIOUSNESS

Consciousness is often used equivocally to refer to several related but dissimilar capacities. In this manuscript, I understand consciousness as the set of phenomenal states experienced simultaneously at any given point in time, which are accessible to the neurocognitive system for use in cognitive processes such as the control of behaviour (Block, 1995; Baars, 2005). Consciousness here is synonymous with phenomenal experience, in that "there is something it is like" to be the conscious organism in question (Nagel, 1974). In its most rudimentary form, consciousness consists of perceptual or sensory experience evoked by external and internal stimuli. These include awareness of one's external surroundings and internal states that have phenomenal qualities (in contrast to those that are not "felt",İ such as blood circulation or digestion). As such, consciousness has both exteroceptive and interoceptive sensory contents.

The literature on consciousness distinguishes between *primary* and *higher-order consciousness,* based on the complexity of its contents. Primary consciousness is that kind of consciousness sometimes equated with the capacity for sensory awareness. For an organism to have primary consciousness, all it needs is the capacity for "direct awareness of the world without further reflection upon that awareness" (Barron and Klein, 2016, p. 4901). Insofar as consciousness attribution is concerned, it is believed that primary consciousness is widespread throughout the animal kingdom. In contrast, higher-order consciousness involves capacities for metacognition that vary in complexity (Seth, 2009). An example of these metacognitive capacities is consciousness of being conscious, i.e., awareness that one is experiencing the conscious states that one is experiencing. Another complex manifestation of higher-order consciousness is a sense of self, wherein the organism recognizes itself as the subject that experiences the experiences it has. It is also believed that more sophisticated forms of higher-order consciousness subserve the ability to mentally "construct past and future scenes" (Seth, 2009, p. 9); a capacity for both short- and long-term memory is thus presupposed in higher-order consciousness.

Another important distinction pertains to the structure of conscious experience. Bayne (2010) distinguishes between a *conscious field* and a *conscious stream.* The former is the cluster of conscious states experienced simultaneously at any single time, while the latter is the series of conscious states experienced over time. It has long been assumed that the "normal" structure of consciousness is that it is *unified,* in that a conscious organism experiences a single conscious field at any given time (Bayne, 2008, 2010). This notion has been putatively challenged by phenomena such as the split-brain syndrome, wherein the corpus callosum is severed (originally to prevent the spread of epilepsy across brain hemispheres). Apparent mostly in experimental settings, the split-brain syndrome involves "information presented in the [right visual field being] unavailable for left-handed grasping behaviour while information presented in the [left visual field is] unavailable for verbal report" (Bayne, 2010, p. 192). Octopuses have appeared to be another challenger to the unity thesis, because of the extensive distribution of their nervous systems and cognitive routines and the considerable autonomy displayed by the highly elaborated peripheral nervous system. While there is accumulating evidence in favor of unity (Mather, 2021), adjudicating whether octopuses experience multiple conscious fields or a single one requires independent investigation outside of present purposes.

The discussion on consciousness will proceed following an overview of what the octopus nervous system is like.

OCTOPUS NERVOUS SYSTEM

Anatomy and Functions

With its 500 million neurons—a number more typical of vertebrates such as dogs—octopuses have the largest nervous systems among invertebrates (Hochner, 2004). The octopus nervous system is highly distributed, and typically divided along anatomical lines into components with considerable functional autonomy. The three main parts of the octopus nervous system are the *brain*, the *optic lobes*, and the highly elaborated *arm*

nervous system. Significantly, the arm nervous system contains three-fifths of the octopus's neurons. Importantly, the brain, optic lobes, and arm nervous system are interconnected by only about 30,000 nerve fibres, suggesting that "much of the processing of motor and sensory information is performed in the peripheral nervous system and the optic lobes" (Hochner, 2012, p. R889).

The central nervous system (CNS) consists of the brain, which with 45–50 million neurons is the smallest component of the nervous system. The brain is responsible for integrating information received from the different parts of the nervous system, as well as high-level "cognitive and executive functions like motor coordination, decisionmaking (*sic*), and learning and memory" (Levy et al., 2017, p. 7). For instance, the brain is responsible for selecting and initiating or terminating a particular behaviour or action, but the details required for realising arm movements are embedded within the arm nervous system (Sumbre et al., 2005, 2006). The brain contains the basal lobes, the highest motor control centres in the octopus. Early on, it was discovered that stimulating different parts of the basal lobe evokes different types of complex movements: "electrical stimulation of the anterior basal lobe. . .produces effective walking movements of the arms, stimulation of the median basal produces swimming, and of the lateral basal changes in colour over the whole skin" (Young, 1971, p. 14). The vertical lobe system, which processes visual memory and are vital for cross-brain transfer of visual information, is also found in the brain. When the vertical lobes are removed, memory transfer is impaired, in direct proportion to the extent of excision.

The paired and lateralised optic lobes are usually considered part of the peripheral nervous system (PNS), but are sometimes regarded as part of the CNS. Between them, the optic lobes have 120–180 million neurons. Each optic lobe is responsible for processing visual information received *via* the ipsilateral eye. Octopus eyes are highly developed and comparable to those of vertebrates. Octopuses are highly visual, especially when it comes to navigation and learning. They have lateralised vision, and are able to use a single eye for perceptual and learning tasks. Signals received *via* one eye are transmitted and processed in its ipsilateral optic lobe, which sends this information further upstream for "cross-brain transfer" (Mather, 2021, p. 408). Tasks learned while using one eye can later be performed with the other eye, although not as accurately or efficiently as with the original one. This is because "cross-brain transfer [of information from one eye] would normally be complete but storage or retrieval would not be as good in the contralateral as in the ipsilateral area of the brain" (Mather, 2021, p. 408).

The most notable—and largest—component of the PNS is the arm nervous system, in which 350 million neurons are distributed equally between the eight functionally and anatomically[1] identical arms. The arms are interconnected to each other and to the brain by a ring of fibres at their bases, often referred to as the *interbrachial commissure*. Within each arm can be found an *axial nerve cord,* four *intramuscular nerve cords, sucker ganglia*, and millions of *sensory receptors* responsive to chemical, tactile, mechanical, and proprioceptive stimulation. The axial nerve cord is a high-level sensory processing and motor control centre, which integrates information from the respective arm with the commands it receives from the CNS (Richter et al., 2015). The intramuscular nerve cords play an important role in the motor control of the arm. They receive proprioceptive information from proprioceptors in the arm, which are embedded outside the *suckers* (Grasso, 2014). On the underside of each arm are numerous highly sensitive suckers arranged in a double row. With thousands of sensory receptors and motor neurons each, suckers are an important source of tactile, chemical, and spatial information (Grasso, 2014). They are used in a great variety of octopus behaviour, such as object manipulation and locomotion, for instance "walking" along a surface (Grasso, 2014). Each sucker is innervated by its own sucker ganglion, which does not communicate directly with other sucker ganglia (Grasso, 2014). Instead, information from one sucker is channelled *via* the axial nerve cord to nearby ones.

Notable Features

There are two features of the octopus nervous system that stand out as being unique and unusual. The first is the brain's inability to support *somatotopic representation* or point-for-point mapping of the body, and the second is the extensive autonomy in sensory processing and motor control of the arm nervous system. These will now be discussed in turn.

Following stimulation experiments to the basal lobes, which are the octopus's highest motor control centres, it was discovered that the octopus brain is incapable of somatotopically representing the body (Zullo et al., 2009). Rather than a somatotopic map, it is likely that what are represented in the octopus brain are motor programmes or functions (Zullo et al., 2009), which can then be consolidated with sensory information sourced from all over the nervous system (Zullo and Hochner, 2011). The absence of a somatotopic map was inferred following findings that direct stimulation to the basal lobes led to identical movements of multiple arms whereas generating movement in a specific arm was not possible, and that the same pattern of movement can be evoked by stimulating different parts within the basal lobes. What these findings imply is that motor commands from the brain are global, and received by multiple, if not all, arms instead of by a particular appendage; it is hypothesised that the brain "generates only one motor command to all arms if they are activated in the same behavioural context" (Levy et al., 2017, p. 12). It may thus be the case that the brain is incapable of proprioceptively distinguishing between individual arms, or if it were, it might not do so robustly. Activation of a single arm for use in a task, such as reaching, would then require extensive participation of the PNS and cannot be accomplished by the brain alone.

Moreover, these findings are in line with the fact that the neural resources of the octopus brain are inadequate to "be able to deal with such a huge number of parameters that would be sufficient to represent its muscular system" (Levy et al., 2017, p. 3). A rigid skeleton would have supplied permanent structures to serve as proprioceptive landmarks that would facilitate somatotopic representation and motor control

[1]An exception is *hectocotylisation,* or a modification for sexual purposes found in the third right arm of male octopuses.

(Wolpert, 1997; Gutfreund et al., 1996). However, octopuses lack a skeleton, and furthermore have soft bodies with arms that are "unsegmented. . .and can deform at any point along their length. Each arm can, at any point along its length, bend in any direction, elongate, shorten, and twist either clockwise or counterclockwise" (Levy et al., 2017, p. 3). The demands of somatotopically representing such a body are exorbitant, and have been proposed by some authors to be beyond of any biological system (Levy et al., 2017). As compensation, the octopus evolved a unique solution to the demands both of monitoring and controlling such a flexible body with countless movement possibilities and processing integrating multi-modal information from its millions upon millions of sensory receptors: the development of a highly elaborated and autonomous PNS.

It has been said of the arm nervous system that it appears to be "in some ways curiously divorced from the rest of the brain and many of the arms' actions are performed without reference to the brain" (Hanlon and Messenger, 1996, p. 15). The extent of such independence from the brain was most dramatically demonstrated in early studies on severed octopus arms (Rowell, 1963). In these studies, it was discovered that touching the suckers evoked the grasping reflex, for up to three hours after the arm had been amputated, thus proving that sucker control was localised within the suckers and their respective ganglia. In the same vein, it was discovered that when pricked, a freshly amputated arm would demonstrate a number of responses identical to those found in intact animals. The first set of findings showed that the suckers would grasp at whatever surfaces they came into contact with, and that the grasping was "stronger than that normally elicited from the intact animal" (Rowell, 1963, p. 259). These findings indicated that the control of suckers is mainly localised within their respective ganglia: the brain may "influence but. . .not specify the detail of ongoing sucker and arm behaviour" (Grasso, 2014, p. 114). The second set of findings demonstrated that when pricked with a needle, the arm generates the following responses: "a local flinching of the skin, due to contraction of the dermal and subdermal muscle layers, a movement of the whole arm withdrawing it from the stimulus, and flexion of the arm and protrusion of the suckers in a way likely to cause them to come into contact with the stimulating object" (Rowell, 1963, p. 259–260).

In the same vein, amputated or neurally isolated arms are able to produce movements in response to electrical or tactile stimulation that are almost identical to those in intact octopuses (Sumbre et al., 2001). These findings demonstrate that "the basic motor programme for voluntary movement is embedded within the neural circuitry of the arm itself" (Sumbre et al., 2001, p. 1845). Thus, whereas selecting and activating motor programmes is the responsibility of the brain, the actual "instructions" for bringing the arm into the required shape are contained within the arm. Since the muscles of octopus arms are arranged *hydrostatically,* wherein contraction in one muscle group produces compensatory lengthening in the others (Kier and Smith, 1985), and do not contain any fixed structures, they have potentially unlimited degrees of freedom (DOFs) of movement. However, in order to simplify motor control, octopuses have evolved a set of *stereotypic motor* programmes (Sumbre et al., 2001, 2006) that are used in the majority of its actions. Thus, rather than formulating commands to bring the arm muscles into the required shape from scratch every time, motor control labour is reduced to orienting the arm correctly and scaling the velocity of the movement (Gutfreund et al., 2006). In addition to simplifying motor control demands, the use of stereotypic motor programmes also dissolves the need for somatotopic representation in the motor centres in the brain.

It has been proposed that octopus arms are capable of somatotopic mapping (Grasso, 2014). To understand how, we must follow Grasso's (2014) deconstruction of the octopus arm into local brachial modules (LBM), which "contain the neural components of each sucker-ganglion/brachial-ganglion pair (i.e., the primary receptors [chemo-, mechano-, and proprioceptive], the motor neurons and the interneurons)" (p. 102). Now, each LBM is provided with sensory information by its respective suckers. Importantly, the rims of the suckers "necessarily form a topologically ordered spatial array" (Grasso, 2014, p. 105). This is because the close double-row arrangement of suckers along the arm entails that each sucker will come into contact with the same object or surface at different locations. Since each sucker transmits information to its own ganglion, this information is "location-specific": each sucker ganglion receives information about a different area of the object or surface in question, which the higher processing centres receiving this input are able to consolidate into a more holistic "picture." Importantly, the activation patterns produced by sucker activity and the movements of the arm that accompany it are "stored and remembered hierarchically *across the network of ganglia* [and] have an ordered spatial arrangement that reflects the attitude of the animal's body and state of the external world as sensed by contact with surfaces" (Grasso, 2014, p. 110). Furthermore, these activation patterns also reflect the temporal sequence in which they occurred. Since the arm moves in order to bring the suckers into contact with the object, some suckers are bound to touch it before others. This entails that the corresponding activation by the LBMs they belong to occurs before others, therefore allowing the arm nervous system to monitor the movement of the arm in question. Additionally, information about the activity of the arm and its suckers may be stored for minutes or possibly even up to an hour and recruited for use in learning, suggesting that the arm nervous system is capable of memory and perhaps even representation (Grasso, 2014). Thus, due to the intrinsic topographical and temporal organisation of information received *via* the sensory and mechanoreceptors within the arm, there is a possibility that in contrast to the brain, "a somatotopic map might be formed by the arm" (Grasso, 2014, p. 115) of that arm.

ATTRIBUTING CONSCIOUSNESS TO OCTOPUSES

Neuroanatomy and Neurophysiology

In the Cambridge Declaration on Consciousness (Low et al., 2012), octopuses were declared part of the list of candidates for consciousness on the basis of having the "neuroanatomical, neurochemical, and neurophysiological substrates of conscious

states along with the capacity to exhibit intentional behaviours" (p. 2). Significantly, octopuses possess brain structures analogous or homologous to those in vertebrates (Shigeno et al., 2018) that are associated with consciousness. Taken together, these structures and their functions in the octopus suggest that if they are, like their vertebrate counterparts, involved in generating CNS-based consciousness, the resulting phenomenal field may be fed by information of multiple modalities from all over the nervous system.

An important structure for consciousness in vertebrates is the thalamus, which receives much of the information that is headed for the cerebral cortex. The thalamus then "transmits this information and. . .receives an even greater number of reciprocal connections back from the cortex" (Blumenfeld, 2016, p. 8). As such, the contents of consciousness are relayed *via* the thalamus (Blumenfeld, 2016). An analogous structure in cephalopods might be the dorsal basal and sub-vertical lobes, as they "receive many input fibres from the entire body *via* direct and indirect pathways from the sub-oesophageal mass, suggesting that it is a relay centre from the "cortically located" frontal and vertical lobes in cephalopod brain (*sic*)," (Shigeno et al., 2018, p. 8). In addition to these structures, the inferior frontal lobe is another potential analogue to the thalamus, being "a major chemo-tactile sensory-motor centre processing information originating from the suckers and arms" that is "is involved in learning and memory recall being part of the so-called chemo-tactile memory system" (Shigeno et al., 2018, p. 8). In its processing functions, it resembles the vertebrate olfactory cortex.

It is believed that the vertical lobe system has deep homology with the cerebral cortex (Shigeno et al., 2018), whose responsibilities include "regulating the overall level of consciousness" (Blumenfeld, 2016, p. 16). The basis of this hypothesis is that certain behaviours, in particular "sleeping, decision-making, discrimination learning and lateralisation of the brain" (Shigeno et al., 2018, p. 9) exhibited by some cephalopods such as octopuses are may be indicators of advanced cognitive capacities. Since in mammals, at least, such behaviours require a cerebral cortex, the presence of an equivalent structure in cephalopods must be inferred. The vertical lobe system is involved in the processing of tactile and visual memories, and when removed "impairs long-term memory for new tasks" (Hochner, 2004, p. 4). In its roles in memory in learning, the vertical lobe is similar to the hippocampus in vertebrates (Hochner, 2004). It has also been discovered in early studies that the vertical lobe "is somehow connected with restraint. . .and [might serve] to introduce into the system the effect of nerve fibres [from the arms and mantle] that signal trauma (pain)" (Young, 1971, p. 244). Furthermore, together with the subvertical lobe, the vertical lobe may "amplify such pain signals, in the sense of putting them into more channels, and to insert them in such a way that they have an appropriate effect on the system" (Young, 1971, p. 244).

Behaviour

The sophisticated and complex behavioural repertoire of octopuses is likewise notable because it is the outcome of *domain-general cognition* (Vitti, 2013). In contrast to *domain-specificity,* wherein cognitive capacities are limited to those that are immediately required to survive within a particular ecological niche, domain-generality recruits multiple cognitive domains and thus produces cognitive abilities and behaviour that are flexible and adaptive within a wide variety of situations. Since domain-generality is facilitated by a centralised organisation of the nervous system, it is typically associated with vertebrates, who have highly centralised neurocognitive systems. Furthermore, the emergence of domain-general cognition is believed to have been influenced by sociality, which demands the capacity for adaptive responses to conspecifics and non-conspecifics alike. As such, it is somewhat surprising that octopuses, with their decentralised and distributed nervous systems, and largely solitary life styles would exhibit such behavioural capacities.

However, although modern octopuses are for the most part solitary, their evolutionary history reveals the heavy ecological demands that would have encouraged the emergence of sophisticated cognition and adaptive behaviour. Known as the Packard scenario (Packard, 1972), after its proponent Andrew Packard, the predominant theory is that due to the internalisation and reduction of the ancestral shell—a feature of all coleoid cephalopods, but none more extensive than in octopuses—octopuses lost the capacity for buoyancy. They consequently sank down to the benthos or sea floor—to this day is their natural habitat—which is rich in ecological diversity. In order to survive, octopuses would have had to learn how to interact adaptively with a large number of fellow benthic species—many of which were vertebrates—predator and prey alike (Borrelli and Fiorito, 2008). These ecological pressures thus set the stage for the development of their cognitive and behavioural sophistication, much of which has attracted the attention of researchers. This section presents a number of examples of octopus behaviour that suggest the presence of consciousness.

Octopuses have demonstrated capacities for learning, with regards to behaviour and discriminatory tasks. For instance, when handling unyielding bivalve prey, they employ different techniques selected through trial and error until they are able to get at the edible portions (Mather, 2008). This stands in contrast to perseverating with an ineffective technique, which suggests lower cognitive flexibility. Octopuses are also known for unpredictability and plasticity, rather than fixed or stereotyped responses, in their avoidance behaviours in the face of stimuli previously experienced as negative (Mather, 2008). Similar to vertebrates, octopuses are capable of associative and reverse associative learning, sensitisation and habituation to stimuli, using multiple cues in visual discrimination tasks, stimulus generalisation, spatial learning, and conditional discrimination. They have also demonstrated capacities to learn about objects not encountered in the wild, in the form of different types of sensory discriminations. They can visually distinguish between orientations, rotations, and mirror images, as well as tactilely discriminate between shapes, curvature, and striation of objects not encountered in the wild (Wells, 1964; Wells and Wells, 1957). Taken together, these learned discriminations suggest that octopuses may be capable of concept formation (Mather, 2008).

Another important capacity that subserves consciousness is memory. Storage and retrieval of information stored in memory decouples the organism from the environment, as this information remains accessible over time, without requiring the presence of the original stimuli or scenario. Memory also allows the mental reconstruction of past scenes, a task held to be achievable only with the involvement of consciousness; furthermore, where a capacity for planning is present, memory provides information that can be recruited for use in mentally constructing future scenarios and formulating actions needed to bring or avoid certain states of affairs. Octopuses are capable of storing short- and long-term memories, the latter of which can stay stable for months (Hochner et al., 2006)—which is remarkable since their typical life spans are 1–2 years.

Octopuses' capacities for memory are also highlighted in their use and occupancy of dens. Denning behaviour is exhibited by many octopus species, wherein a hole is dug in the seabed or any other soft substrate, and used as a residence for several days to a few weeks. In some cases, octopuses collect stones and arrange them around the opening of the den. Octopuses usually capture prey by going on hunting trips that can last up to several hours and cover large distances, after which they return to the den with the prey to eat. Significantly, they do not use fixed or predictable routes when leaving and returning to the den (Mather, 1991). Furthermore, it has been discovered that octopuses use prominent physical features of the environment as navigational landmarks (Mather, 1991; Hvorecny et al., 2007). In addition to demonstrating the use of memory, denning behaviour is further suggestive of a number of advanced cognitive capacities predominantly observed in vertebrates. Among these are the ability to form mental maps of areas surrounding their dens (Hanlon and Messenger, 1996), the capacity for concept formation manifested as being able to recognize a given feature of the environment from different angles, and conditional discrimination or the ability to "discriminate between potential cues [present in the environment] and show context (condition) sensitivity" (Hvorecny et al., 2007, p. 449). In the context of navigating using environmental landmarks, conditional discrimination is expressed as identifying a certain feature as distinct from similar ones and determining its "significance" in the given context.

Octopuses may also have a sense of self, rudimentary manifestations of which include awareness of one's own physical boundaries that demarcate one from the external world (see also Merker (2005), Godfrey-Smith (2013)), and the capacity to distinguish between oneself and another organism. It has been discovered that octopuses are able to distinguish between themselves and conspecifics through the use of chemoreception (Nesher et al., 2014) and vision (Tricarico et al., 2011). For instance, when presented with their own severed arms and those of conspecifics, octopuses exhibited differing behavioural responses, mediated by chemoreception. The test subjects were more likely to treat the arms of conspecifics as food objects than they did their own (Nesher et al., 2014). Octopuses are also able to recognize individual conspecifics, inferred from the increased tendency for aggressive behaviour toward other octopuses they had not previously encountered (Tricarico et al., 2011). Furthermore, octopuses' individual recognition capacities also extend to humans. In a study by Anderson et al. (2010), identically dressed human handlers who would repeat respectively assigned behaviours regularly approached the octopuses over several days. Some of the handlers consistently offered the octopuses food, while the others would consistently poke them with a brush. Eventually, the octopuses exhibited markedly dissimilar behaviour toward the humans depending on whether they were food-bearing or obnoxious: whereas they approached the former, they tended to be more aggressive or avoidant toward the latter.

It may also be the case that if present, the sense of self in octopuses may go beyond simple self-other demarcation. This is suggested by observations that some octopus species, such as the mimic octopus *(Thaumoctopus mimicus)* rearrange their body outline, colouration, and texture and copy the locomotion techniques of non-conspecifics in order to imitate them (Hanlon and Messenger, 1996; Norman et al., 2001; Hanlon, 2007). This is usually done in potentially hazardous situations. For instance, when swimming across sand plains that offer little opportunities for hiding, octopuses may mimic flounders, which are less appealing to possible predators than octopuses are (Hanlon, 2007). When swimming in predator infested-waters, octopuses have also been observed to copy the posture and striped colouration of venomous lionfish in order to increase chances of safe passage (Norman et al., 2001). Octopuses are also known to pretend to be drifting seaweed (Hanlon and Messenger, 1996), especially when higher up in the water column. One technique used in this task is known as countershading, wherein certain parts of the body are darkened in order to resemble shadows cast by down-welling light. Together, these sophisticated forms of *crypsis* or disguise behaviour suggest that octopuses may be capable of awareness about how they appear from a third-person perspective, a capacity said to be dependent on consciousness and a sense of self.

Finally, the ability to sleep, which octopuses possess (Brown et al., 2006; Meisel et al., 2011; de Souza Medeiros et al., 2021), is also suggestive of consciousness. Along with attention and alertness, the awake state is typically regarded as an indication of a relatively high *level* of consciousness (in the sense of the intensity of conscious awareness), as it is "necessary for any meaningful responses to occur" (Blumenfeld, 2016, p. 4). In contrast, states such as sleep or coma are indicative of a low level of conscious awareness and arousal. As such, the capacity for sleep implies that an organism is able to alternate between states with high and low levels of consciousness (Siclari and Tononi, 2016). However, more detailed characteristics of consciousness in the organism in question are difficult to infer solely from the capacity to sleep.

These behaviours are among those suggestive of consciousness in octopuses. Importantly, they are capacities of the *intact* octopus, i.e., they emerge as the result of the complex interaction between the components of the nervous system. Consequently, investigations into consciousness in octopuses are based on a construal of the animal as a *coherent agent* (Godfrey-Smith, 2020), whose complex behaviour is the outcome of profound embodiment (Hochner, 2013) that evolved as a unique solution to the challenge of controlling a flexible body with immense

sensory processing demands. Now, although we are unlikely to ever have complete knowledge about what it is like to be an octopus, attempting to understand consciousness in such creatures requires that we take a closer view at its arm nervous system, whose participation in cognition and behaviour is vital and indispensable.

ARM-BASED CONSCIOUSNESS?

We cannot say for certain, given knowledge about an animal's nervous system or parts of it, what any conscious experience that arises from it might be like (Nagel, 1974; Chalmers, 1995). Although we can speculate what kinds of physiological states enter into a given creature's consciousness (Morsella, 2005), this is difficult to ascertain from a third-person perspective. Thus, I will sidestep the task proper of consciousness attribution for now, and instead provide principled reasons for surmising that consciousness might exist in the arm nervous system.

The proposal that the octopus arm may be able to generate and support a local conscious field is motivated by the studies of Rowell (1963) and later on Sumbre et al. (2001) on amputated appendages, which demonstrate capacities for sensation and movement (or rudimentary action). If present, this arm-based consciousness would likely be *primary* consciousness, for which "a direct awareness of the world" (Barron and Klein, 2016, p. 4901) suffices. In the same vein, Peter Godfrey-Smith writes that minds (understood as equivalent to a conscious field) "have what we might call *characteristic interfaces*. . .that connect them with external objects and conditions. Sensing and action are the interfaces, and these mark the boundaries of a mind" (Godfrey-Smith, 2020, p. 290). Consequently, demarcating a unit that potentially generates a conscious field involves identifying constituent substrates that are responsible for sensation and action. As it is, the octopus arm is a structure that lends itself somewhat cleanly to such a demarcation task (at least in comparison to bisected human brains exhibiting the split-brain syndrome).

Noting that neither complexity of a process nor the need to integrate information from various sources in the nervous system is sufficient to guarantee a conscious field, Ezequiel Morsella (2005) proposed that the states that do enter into consciousness or are accompanied by phenomenal experience are those that are involved in the control of skeletal muscle. The reason for this is that the effectors can get in each other's way when motor commands or action policies are not harmonised. He speculates that "conscious processes. . .mediate large-scale skeletomotor conflicts caused by structures in the brain with different agendas [and] behavioural tendencies. . ." (Morsella, 2005, p. 1010). As such, "phenomenal states could be considered as one of the mechanisms solving the problem of integrating processes in a largely parallel brain that must satisfy the demands of a skeletomotor system that can often express actions and goals only one at a time." (Morsella, 2005, p. 1010). In other words, consciousness can help ensure that complex behaviour requiring the coordination of multiple effectors is carried out coherently, in part by interoceptively monitoring the effectors as they proceed with their movements. Likewise, Merker (2005) also suggests that the evolutionary emergence of consciousness was influenced by the need to distinguish between sensations caused by externally generated causes and internally generated ones, known as the *reafference problem*. Such a distinction is important, as it is prerequisite to determining whether a behavioural response to such signals is necessary or not, or what kind of response is warranted.

Importantly, these accounts, particularly Morsella's, are largely vertebrate-based. However, assuming that the reason states involved in the control of skeletal muscles are conscious is not because they control skeletal muscles *per se*, but because of the need to ensure that effectors with limited motor capacities need to be harmonised in their movements, the same principle may apply to octopus arms. (If anything, such roles of consciousness might even be more beneficial or adaptive in octopuses given the virtually unlimited motor opportunities available to their arms). Consequently, these accounts can be recruited to help establish why, if ever, conscious experience evolved in octopus arms. The hydrostatic nature of octopus arm muscle entails that the stiffening of certain groups "provides skeletal support against which muscle contractions generate the movements" (Levy et al., 2017, p. 5). Thus, in a sense, the arm muscles are able to function as a sort of pseudo-skeleton that can be dissolved and reconstructed in different ways anywhere on the arm. Although there are mechanisms that prevent octopus arms with interfering with each other, such as chemical mechanisms in the skin that prevent the suckers of an arm from grasping another of the same animal's other appendages (Nesher et al., 2014), and potential existence of gating mechanisms that direct arm extension commands to certain arms and not others (Zullo et al., 2009), these do not rule out the possibility of the presence of a conscious field in the octopus arm.

Although there is reason to believe that intact octopuses experience a single conscious field (Mather, 2021), the question remains as to where phenomenal experience in these animals is generated. In familiar models of consciousness, it is mostly the case that the CNS—particularly the brain—is the sole organ complex enough to be capable of generating a conscious field. This is not so in octopuses, whose arm nervous systems may be sophisticated enough to give rise to phenomenal consciousness, albeit rudimentary. If present, arm-based consciousness would consist of capacities for direct awareness about the world (Barron and Klein, 2016) and motor responses to active stimulation. Importantly, acknowledging that this is even a possibility entails a commitment to the view that consciousness comes in a spectrum of complexity, depending on its contents and the cognitive capacities it may engender or enable, ranging from the very simple to the highly sophisticated.

What I have suggested is that individual octopus arms may generated respective conscious fields, such that in an intact and anatomically normal octopus there may be eight of them. However, due to the profound interconnectedness of the arms into the network that is the arm nervous system, these fields may be experienced not disjointedly as a single field, which is then further incorporated into the conscious field generated by the brain. Now, whether the octopus experiences one single unified field or multiple distinct ones depends on how well they are

bound together (O'Brien and Opie, 1998; Bayne, 2010) or fused into a conjoint phenomenal state.

If indeed octopuses experience a single, unified consciousness, then arm-based consciousness resembles the putative "two minds" of the split-brain syndrome: the multiplicity of conscious fields can only be manifested under conditions very different from the organism's day-to-day experiences. In split-brain cases, this condition would be a specifically designed experimental setup; in octopuses it would be detachment of the arm. Without being subsumed under the octopus's broader conscious field, the conscious field of the detached arm would simply be the conscious field of a detached octopus arm.

CONCLUDING REMARKS

Being largely (and admittedly) speculative, the purpose of this manuscript is to motivate interest and set the stage for future research on the possibility that brains may not be the sole neural structure capable of generating consciousness. This is an important point in the study of animal minds: if we are to have a more comprehensive understanding of different types of creature consciousness, particularly those in invertebrates, we need to go beyond vertebrate-based assumptions about phenomenal experience. Among these assumptions are the notions that consciousness is by necessity unified, that there is only one conscious field per organism, and that only the CNS can generate conscious fields. There is no better case study in these possibilities than octopus arms and their idiosyncratic capacities.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

ACKNOWLEDGMENTS

Many thanks to Alice Laciny and Ivan Gonzalez-Cabrera. I am also grateful to the University of Konstanz for support for this publication.

REFERENCES

Anderson, R. C., Mather, J. A., Monette, M. Q, Zimsen, S. R. (2010). Octopuses (*Enteroctopus Dofleini*) Recognize Individual Humans. *J. Appl. Anim. Welf. Sci.* 13, 261–272. doi: 10.1080/10888705.2010.483892

Baars, B. J. (2005). "Global Workspace Theory of Consciousness: toward a Cognitive Neuroscience of Human Experience," in *Progress in Brain Research*, ed. S. Laureys (Amsterdam: Elsevier), 45–53. doi: 10.1016/S0079-6123(05)50004-9

Barron, A. B., and Klein, C. (2016). What Insects can Tell Us about the Origins of Consciousness. *Proc. Natl. Acad. Sci. U. S. A.* 113, 4900–4908. doi: 10.1073/pnas.1520084113

Bayne, T. (2008). The Unity of Consciousness and the Split-Brain Syndrome. *J. Philos.* 105, 277–300. doi: 10.5840/jphil2008105638

Bayne, T. (2010). *The Unity of Consciousness*. Oxford: Oxford University Press.

Block, N. (1995). On a Confusion about a Function of Consciousness. *Behav. Brain Sci.* 18, 227–287. doi: 10.1017/s0140525x00038188

Blumenfeld, H. (2016). "Neuroanatomical Basis of Consciousness," in *The Neurology of Consciousness*, 2nd Edn, eds S. Laureys, O. Gosseries, and G. Tononi (San Diego: Academic Press), 3–29. doi: 10.1016/b978-0-12-800948-2.00001-7

Borrelli, L., and Fiorito, G. (2008). "Behavioural Analysis of Learning and Memory in Cephalopods," in *Learning and Memory: a Comprehensive Reference*, ed. H. J. Byrne (Amsterdam: Elsevier), 605–627. doi: 10.1016/b978-012370509-9.00069-3

Brown, E. R., Piscopo, S., De Stefano, R., and Giuditta, A. (2006). Brain and Behavioural Evidence for Rest-Activity Cycles in *Octopus vulgaris*. *Behav. Brain Res.* 172, 355–359. doi: 10.1016/j.bbr.2006.05.009

Chalmers, D. J. (1995). Facing Up to the Problem of Consciousness. *J. Conscious. Stud.* 2, 200–219.

de Souza Medeiros, S. L., Matias de Paiva, M. M., Lopes, P. H., Blanco, W., Dantas de Lima, F., Cirne de Oliveira, J. B., et al. (2021). Cyclic Alternation of Quiet and Active Sleep States in the *Octopus*. *Iscience* 24, 1–19. doi: 10.1016/j.isci.2021.102223

Godfrey-Smith, P. (2013). Cephalopods and the Evolution of the Mind. *Pac. Conserv. Biol.* 19, 4–9. doi: 10.1071/pc130004

Godfrey-Smith, P. (2020). Integration, Lateralization, and Animal Experience. *Mind Lang.* 36, 285–296. doi: 10.1111/mila.12323

Grasso, F. W. (2014). "The *Octopus* with Two Brains: how are Distributed and Central Representations Integrated in the *Octopus* Central Nervous System?," in *Cephalopod Cognition*, eds A.-S. Darmaillacq, L. Dickel, and J. Mather (Cambridge: Cambridge University Press), 94–122. doi: 10.1017/cbo9781139058964.008

Gutfreund, Y., Flash, T., Yarom, Y., Fiorito, G., Segev, I., and Hochner, B. (1996). Organization of octopus arm movements: a model system for studying the control of flexible arms. *J. Neurosci.* 16, 7297–7307.

Gutfreund, Y., Matzner, H., Flash, T., and Hochner, B. (2006). Patterns of Motor Activity in the Isolated Nerve Cord of the *Octopus* Arm. *Biol. Bull.* 211, 212–222. doi: 10.2307/4134544

Hanlon, R. (2007). Cephalopod Dynamic Camouflage. *Curr. Biol.* 17, R400–R404.

Hanlon, R. T., and Messenger, J. B. (1996). *Cephalopod Behaviour*. Cambridge: Cambridge University Press.

Hochner, B. (2004). "*Octopus* Nervous System," in *Encyclopedia of Neuroscience*, 3rd Edn, eds G. Adelman and B. H. Smith (Amsterdam: Elsevier).

Hochner, B. (2012). An Embodied View of *Octopus* Neurobiology. *Curr. Biol.* 22, R887–R892.

Hochner, B. (2013). How Nervous Systems Evolve in Relation to their Embodiment: what we can Learn from Octopuses and Other Molluscs. *Brain Behav. Evol.* 82, 19–30. doi: 10.1159/000353419

Hochner, B., Shomrat, T., and Fiorito, G. (2006). The *Octopus*: a Model for a Comparative Analysis of the Evolution of Learning and Memory Mechanisms. *Biol. Bull.* 210, 308–317.

Hvorecny, L. M., Grudowski, J. L., Blakeslee, C. J., Simmons, T. L., Roy, P. R., Brooks, J. A., et al. (2007). *Octopuses (Octopus Bimaculoides)* and cuttlefishes *(Sepia Pharaonis, S. Officinalis)* can Conditionally Discriminate. *Anim. Cogn.* 10, 449–459. doi: 10.1007/s10071-007-0085-4

Kier, W. M., and Smith, K. K. (1985). Tongues, Tentacles and Trunks: the Biomechanics of Movement in Muscular-Hydrostats. *Zool. J. Linn. Soc.* 83, 307–324.

Levy, G., Nesher, N., Zullo, L., and Hochner, B. (2017). "Motor Control in Soft-Bodied Animals: the *Octopus*," in *The Oxford Handbook of Invertebrate Neurobiology*, ed. J. H. Byrne (Oxford: Oxford Handbooks Online), 1–27.

Low, P., Panksepp, J., Reiss, D., Edelman, D., van Swinderen, B., and Koch, C. (2012). "The Cambridge declaration on consciousness," in *Proceedings of the Francis Crick Memorial Conference on Consciousness in Human and non-Human Animals*, (Cambridge: University of Cambridge). doi: 10.1016/b978-0-323-07909-9.09001-2

Mather, J. (2021). The Case for *Octopus* Consciousness: unity. *Neurosci* 2, 405–415. doi: 10.3390/neurosci2040030

Mather, J. A. (1991). Navigation by Spatial Memory and use of Visual Landmarks in Octopuses. *J. Comp. Psychol.* 168, 491–497. doi: 10.1007/bf00199609

Mather, J. A. (2008). Cephalopod Consciousness: behavioural Evidence. *Conscious. Cogn.* 17, 37–48. doi: 10.1016/j.concog.2006.11.006

Meisel, D. V., Byrne, R. A., Mather, J. A., and Kuba, M. (2011). Behavioral Sleep in *Octopus vulgaris*. *Vie Et Milieu* 61, 185–190.

Merker, B. (2005). The Liabilities of Mobility: a Selection Pressure for the Transition to Consciousness in Animal Evolution. *Conscious. Cogn.* 14, 89–114. doi: 10.1016/S1053-8100(03)00002-3

Morsella, E. (2005). The Function of Phenomenal States: supramodular Interaction Theory. *Psychol. Rev.* 112, 1000–1021. doi: 10.1037/0033-295X.112.4.1000

Nagel, T. (1974). What is it Like to be a Bat? *Philos. Rev.* 83, 435–450. doi: 10.1111/1468-5930.00141

Nesher, N., Levy, G., Grasso, F. W., and Hochner, B. (2014). Self-Recognition Mechanism between Skin and Suckers Prevents *Octopus* Arms from Interfering with each Other. *Curr. Biol.* 24, 1271–1275. doi: 10.1016/j.cub.2014.04.024

Norman, M. D., Finn, J., and Tregenza, T. (2001). Dynamic Mimicry in an Indo-Malayan *Octopus*. *Proc. R. Soc. Lond. B Biol. Sci.* 268, 1755–1758.

O'Brien, G., and Opie, J. (1998). The Disunity of Consciousness. *Australas. J. Philos.* 76, 378–395.

Packard, A. (1972). Cephalopods and Fish: the Limits of Convergence. *Biol. Rev.* 47, 241–307.

Richter, J. N., Hochner, B., and Kuba, M. J. (2015). *Octopus* Arm Movements Under Constrained Conditions: adaptation, Modification and Plasticity of Motor Primitives. *J. Exp. Biol.* 218, 1069–1076. doi: 10.1242/jeb.115915

Rowell, C. H. F. (1963). Excitatory and Inhibitory Pathways in the Arm of *Octopus*. *J. Exp. Biol.* 40, 257–270.

Seth, A. (2009). Functions of Consciousness. *PsyArXiv* [Preprint]. doi: 10.31234/osf.io/wybkp

Shigeno, S., Andrews, P. L. R., Ponte, G., and Fiorito, G. (2018). Cephalopod Brains: an Overview of Current Knowledge to Facilitate Comparison with Vertebrates. *Front. Physiol.* 9:952. doi: 10.3389/fphys.2018.00952

Siclari, F., and Tononi, G. (2016). "Sleep and Dreaming," in *The Neurology of Consciousness*, 2nd Edn, eds S. Laureys, O. Gosseries, and G. Tononi (San Diego: Academic Press), 107–128.

Sumbre, G., Fiorito, G., Flash, T., and Hochner, B. (2005). Motor Control of Flexible *Octopus* Arms. *Nature* 433, 595–596.

Sumbre, G., Fiorito, G., Flash, T., and Hochner, B. (2006). Octopuses use a Human-Like Strategy to Control Precise Point-to-Point Arm Movements. *Curr. Biol.* 16, 767–772. doi: 10.1016/j.cub.2006.02.069

Sumbre, G., Gutfreund, Y., Fiorito, G., Flash, T., and Hochner, B. (2001). Control of *Octopus* Arm Extension by a Peripheral Motor Program. *Science* 293, 1845–1848. doi: 10.1126/science.1060976

Tricarico, E., Borrelli, L., Gherardi, F., and Fiorito, G. (2011). I Know My Neighbour: individual Recognition in *Octopus vulgaris*. *PLoS One* 6:e18710. doi: 10.1371/journal.pone.0018710

Vitti, J. J. (2013). Cephalopod Cognition in an Evolutionary Context: implications for Ethology. *Biosemiotics* 6, 393–401.

Wells, M. J. (1964). Tactile Discrimination of Surface Curvature and Shape by the *Octopus*. *J. Exp. Biol.* 41, 433–445.

Wells, M. J., and Wells, J. (1957). The Function of the Brain of Octopus in Tactile Discrimination. *Journal of Experimental Biology* 34, 131–142.

Wolpert, D. M. (1997). Computational Approaches to Motor Control. *Trends Cogn. Sci.* 1, 209–216.

Young, J. Z. (1971). *Anatomy of the Nervous System of Octopus Vulgaris*. Oxford: Oxford University Press.

Zullo, L., and Hochner, B. (2011). A New Perspective on the Organization of an Invertebrate Brain. *Commun. Integr. Biol.* 4, 26–29. doi: 10.4161/cib.4.1.13804

Zullo, L., Sumbre, G., Agnisola, C., Flash, T., and Hochner, B. (2009). Nonsomatotopic Organization of the Higher Motor Centers in *Octopus*. *Curr. Biol.* 19, 1632–1636. doi: 10.1016/j.cub.2009.07.067

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Systems Neuroscience

HYPOTHESIS AND THEORY
published: 24 March 2022
doi: 10.3389/fnsys.2022.768201

Technological Approach to Mind Everywhere: An Experimentally-Grounded Framework for Understanding Diverse Bodies and Minds

Michael Levin[1,2]*

[1] Allen Discovery Center at Tufts University, Medford, MA, United States, [2] Wyss Institute for Biologically Inspired Engineering at Harvard University, Cambridge, MA, United States

Synthetic biology and bioengineering provide the opportunity to create novel embodied cognitive systems (otherwise known as minds) in a very wide variety of chimeric architectures combining evolved and designed material and software. These advances are disrupting familiar concepts in the philosophy of mind, and require new ways of thinking about and comparing truly diverse intelligences, whose composition and origin are not like any of the available natural model species. In this Perspective, I introduce TAME—Technological Approach to Mind Everywhere—a framework for understanding and manipulating cognition in unconventional substrates. TAME formalizes a non-binary (continuous), empirically-based approach to strongly embodied agency. TAME provides a natural way to think about animal sentience as an instance of collective intelligence of cell groups, arising from dynamics that manifest in similar ways in numerous other substrates. When applied to regenerating/developmental systems, TAME suggests a perspective on morphogenesis as an example of basal cognition. The deep symmetry between problem-solving in anatomical, physiological, transcriptional, and 3D (traditional behavioral) spaces drives specific hypotheses by which cognitive capacities can increase during evolution. An important medium exploited by evolution for joining active subunits into greater agents is developmental bioelectricity, implemented by pre-neural use of ion channels and gap junctions to scale up cell-level feedback loops into anatomical homeostasis. This architecture of multi-scale competency of biological systems has important implications for plasticity of bodies and minds, greatly potentiating evolvability. Considering classical and recent data from the perspectives of computational science, evolutionary biology, and basal cognition, reveals a rich research program with many implications for cognitive science, evolutionary biology, regenerative medicine, and artificial intelligence.

Keywords: regeneration, basal cognition, bioelectricity, gap junctions, synthetic morphology, bioengineering

OPEN ACCESS

Edited by:
Eva Jablonka,
Tel Aviv University, Israel

Reviewed by:
Lars Chittka,
Queen Mary University of London, United Kingdom
Louis Neal Irwin,
The University of Texas at El Paso, United States
Patrick McGivern,
University of Wollongong, Australia

***Correspondence:**
Michael Levin
michael.levin@tufts.edu

Received: *31 August 2021*
Accepted: *24 January 2022*
Published: *24 March 2022*

Citation:
Levin M (2022) Technological Approach to Mind Everywhere: An Experimentally-Grounded Framework for Understanding Diverse Bodies and Minds. Front. Syst. Neurosci. 16:768201. doi: 10.3389/fnsys.2022.768201

INTRODUCTION

All known cognitive agents are collective intelligences, because we are all made of parts; biological agents in particular are not just structurally modular, but made of parts that are themselves agents in important ways. There is no truly monadic, indivisible yet cognitive being: all known minds reside in physical systems composed of components of various complexity and active behavior. However, as human adults, our primary experience is that of a centralized, coherent Self which controls events in a top-down manner. That is also how we formulate models of learning ("the *rat* learned X"), moral responsibility, decision-making, and valence: at the center is a subject which has agency, serves as the locus of rewards and punishments, possesses (as a single functional unit) memories, exhibits preferences, and takes actions. And yet, under the hood, we find collections of cells which follow low-level rules *via* distributed, parallel functionality and give rise to emergent system-level dynamics. Much as single celled organisms transitioned to multicellularity during evolution, the single cells of an embryo construct *de novo,* and then operate, a unified Self during a single agent's lifetime. The compound agent supports memories, goals, and cognition that belongs to that Self and not to any of the parts alone. Thus, one of the most profound and far-reaching questions is that of scaling and unification: how do the activities of competent, lower-level agents give rise to a multiscale holobiont that is truly more than the sum of its parts? And, given the myriad of ways that parts can be assembled and relate to each other, is it possible to define ways in which truly diverse intelligences can be recognized, compared, and understood?

Here, I develop a framework to drive new theory and experiment in biology, cognition, evolution, and biotechnology from a multi-scale perspective on the nature and scaling of the cognitive Self. An important part of this research program is the need to encompass beings beyond the familiar conventional, evolved, static model animals with brains. The gaps in existing frameworks, and thus opportunities for fundamental advances, are revealed by a focus on plasticity of existing forms, and the functional diversity enabled by chimeric bioengineering. To illustrate how this framework can be applied to unconventional substrates, I explore a deep symmetry between behavior and morphogenesis, deriving hypotheses for dynamics that up- and down-scale Selves within developmental and phylogenetic timeframes, and at the same time strongly impact the speed of the evolutionary process itself (Dukas, 1998). I attempt to show how anatomical homeostasis can be viewed as the result of the behavior of the swarm intelligence of cells, and provides a rich example of how an inclusive, forward-looking technological framework can connect philosophical questions with specific empirical research programs.

The philosophical context for the following perspective is summarized in **Table 1** (see also **Glossary**), and links tightly to the field of basal cognition (Birch et al., 2020) *via* a fundamentally gradualist approach. It should be noted that the specific proposals for biological mechanisms that scale functional capacity are synergistic with, but not linearly dependent on, this conceptual basis. The hypotheses about how bioelectric networks scale cell computation into anatomical homeostasis, and the evolutionary dynamics of multi-scale competency, can be explored without accepting the "minds everywhere" commitments of the framework. However, together they form a coherent lens onto the life sciences which helps generate testable new hypotheses and integrate data from several subfields.

TABLE 1 | The core tenets of TAME.

- Continuum of cognitive capacities—no binary categories, no bright line separating true cognition from "just physics," as is clear from evolutionary process and ability to bioengineer chimeras between any two "natural kinds."
- Mature frameworks must apply to truly diverse intelligences—beyond the examples from Earth's phylogenetic tree based on brains, we must be able to consider and compare agents across the option space of designed and evolved combinations of living, non-living, and software components at all scales.
- Selves exist across a continuum of persuadability, and it is an empirical question as to where on this axis any given system lies (revealed by the ratio of prediction and control vs. effort and knowledge that needs to be input, for any given way of relating to that system).
- Selves are not fixed, permanent agents—their substrate can remodel radically during their lifetime; the owner of memories and preferences, and the subject that interprets rewards and punishments, is malleable and plastic.
- The core of being a Self is the ability to pursue goals. Selves can be nested and overlapping, cooperating and competing both laterally and across levels. Each higher-level self deforms the option space for the lower level Selves, enabling them to follow energy minimization to achieve outcomes that look inevitable and simple at one scale, while serving intelligent goals at a higher scale.
- Intelligence is the degree of competency of navigating any space (not just the familiar 3D space of motility), including morphospace, transcriptional space, physiological space, etc., toward desirable regions, while avoiding being trapped in local minima. Estimates of intelligence of any system are observer-dependent, and say as much about the observer and their limitations as they do about the system itself.

For the purposes of this paper, "cognition" refers not only to complex, self-reflexive advanced cognition or metacognition, but is used in the less conservative sense that recognizes many diverse capacities for learning from experience (Ginsburg and Jablonka, 2021), adaptive responsiveness, self-direction, decision-making in light of preferences, problem-solving, active probing of their environment, and action at different levels of sophistication in conventional (evolved) life forms as well as bioengineered ones (Rosenblueth et al., 1943; Lyon, 2006; Bayne et al., 2019; Levin et al., 2021; Lyon et al., 2021; **Figure 1**). For our purposes, cognition refers to the functional computations that take place between perception and action, which allow the agent to span a wider range of time (*via* memory and predictive capacity, however much it may have) than its immediate *now*, which enable it to generalize and infer patterns from instances of stimuli—precursors to more advanced forms of recombining concepts, language, and logic.

The framework, TAME—Technological Approach to Mind Everywhere—adopts a practical, constructive engineering perspective on the optimal place for a given system on the continuum of cognitive sophistication. This gives rise to an axis of *persuadability* (**Figure 2**), which is closely related to the Intentional Stance (Dennett, 1987) but made more explicit in terms of functional engineering approaches needed to implement prediction and control in practice. Persuadability refers to the

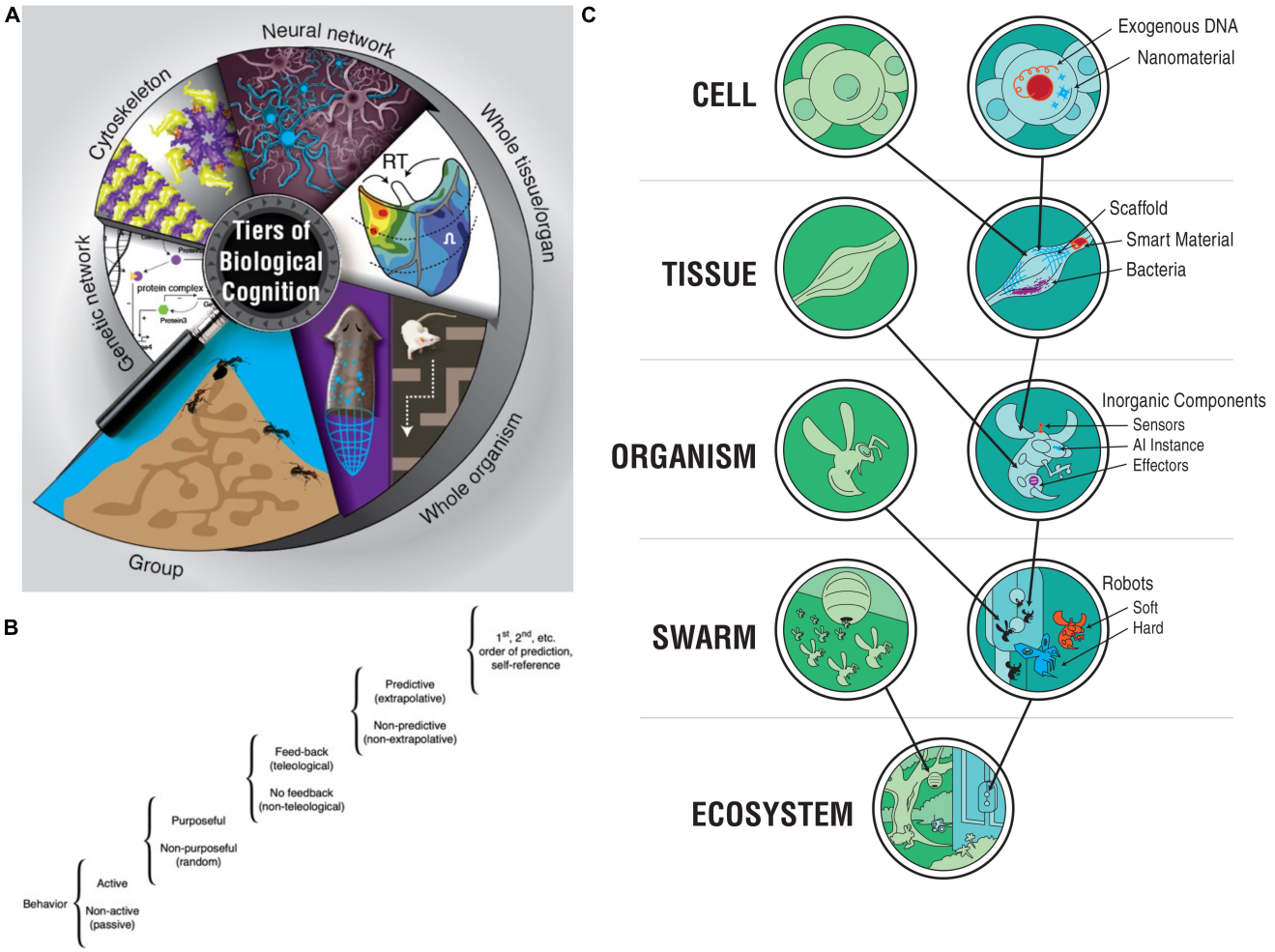


FIGURE 1 | Diverse, multiscale intelligence. **(A)** Biology is organized in a multi-scale, nested architecture of molecular pathways. **(B)** These are not merely structural, but also computational: each level of this holarchy contains subsystems which exhibit some degree of problem-solving (i.e., intelligent) activity, on a continuum such as the one proposed by Rosenblueth et al. (1943). **(C)** At each layer of a given biosystem, novel components can be introduced of either biological or engineered origin, resulting in chimeric forms that have novel bodies and novel cognitive systems distinct from the typical model species on the Earth's phylogenetic lineage. Images in panels **(A,C)** by Jeremy Guay of Peregrine Creative. Image in panel **(B)** was created after Rosenblueth et al. (1943).

type of conceptual and practical tools that are optimal to rationally modify a given system's behavior. The origin story (designed vs. evolved), composition, and other aspects are not definitive guides to the correct level of agency for a living or non-living system. Instead, one must perform experiments to see which kind of intervention strategy provides the most efficient prediction and control (thus, one aim should be generalizing the human-focused Turing Test and other IQ metrics into a broader agency detection toolkit, which perhaps could itself be implemented by a useful algorithm).

Our capacity to find new ways to understand and manipulate complex systems is strongly related to how we categorize agency in our world. Newton didn't invent two terms—gravity (for terrestrial objects falling) and perhaps *shmavity* (for the moon)—because it would have lost out on the much more powerful unification. TAME proposes a conceptual unification that would facilitate porting of tools across disciplines and model systems. We should avoid quotes around mental terms because there is no absolute, binary distinction between *it knows* and *it "knows"*—only a difference in the degree to which a model will be useful that incorporates such components.

Given this perspective, below I develop hypotheses about invariants that unify otherwise disparate-seeming problems, such as morphogenesis, behavior, and physiological allostasis. I take goals (in the cybernetic sense) and stressors (as a system-level result of distance from one's goals) as key invariants which allow us to study and compare agents in truly diverse embodiments. The processes which scale goals and stressors form a positive feedback loop with modularity, thus both arising from, and potentiating the power of, evolution. These hypotheses suggest a specific way to understand the scaling of cognitive capacity through evolution, make interesting predictions, and suggest novel experimental work. They also provide ways to think about the impending expansion of the "space of possible bodies and

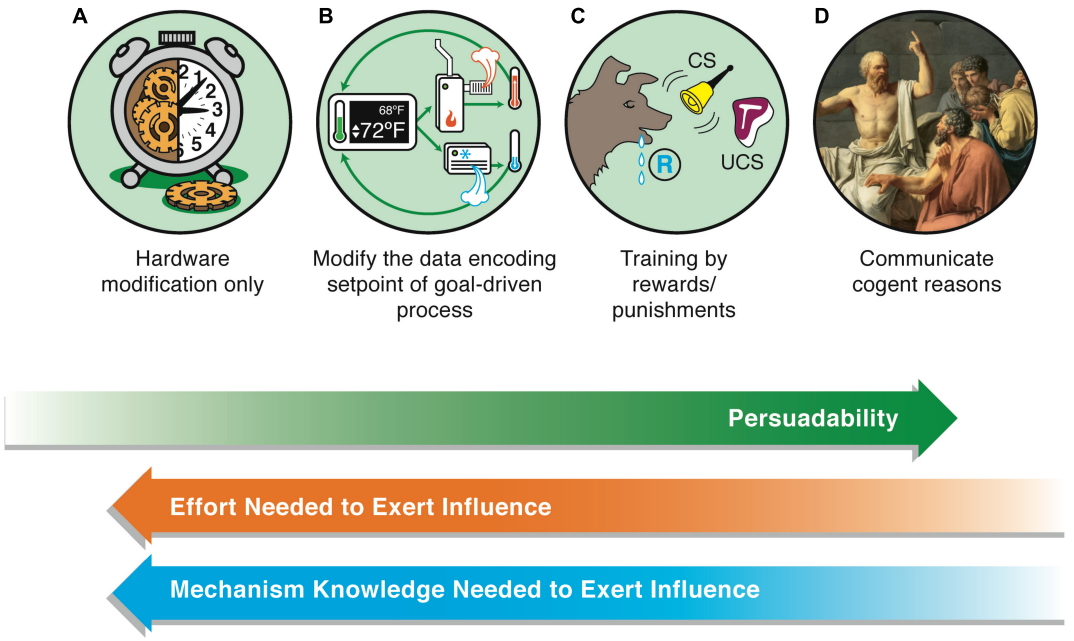


FIGURE 2 | The axis of persuadability. A proposed way to visualize a continuum of agency, which frames the problem in a way that is testable and drives empirical progress, is *via* an "axis of persuadability": to what level of control (ranging from brute force micromanagement to persuasion by rational argument) is any given system amenable, given the sophistication of its cognitive apparatus? Here are shown only a few representative waypoints. On the far left are the simplest physical systems, e.g., mechanical clocks **(A)**. These cannot be persuaded, argued with, or even rewarded/punished—only physical hardware-level "rewiring" is possible if one wants to change their behavior. On the far right **(D)** are human beings (and perhaps others to be discovered) whose behavior can be radically changed by a communication that encodes a rational argument that changes the motivation, planning, values, and commitment of the agent receiving this. Between these extremes lies a rich panoply of intermediate agents, such as simple homeostatic circuits **(B)** which have setpoints encoding goal states, and more complex systems such as animals which can be controlled by signals, stimuli, training, etc., **(C)**. They can have some degree of plasticity, memory (change of future behavior caused by past events), various types of simple or complex learning, anticipation/prediction, etc. Modern "machines" are increasingly occupying right-ward positions on this continuum (Bongard and Levin, 2021). Some may have preferences, which avails the experimenter of the technique of rewards and punishments—a more sophisticated control method than rewiring, but not as sophisticated as persuasion (the latter requires the system to be a logical agent, able to comprehend and be moved by arguments, not merely triggered by signals). Examples of transitions include turning the sensors of state outward, to include others' stress as part of one's action policies, and eventually the meta-goal of committing to enhance one's agency, intelligence, or compassion (increase the scope of goals one can pursue). A more negative example is becoming sophisticated enough to be susceptible to a "thought that breaks the thinker" (e.g., existential or skeptical arguments that can make one depressed or even suicidal, Gödel paradoxes, etc.)—massive changes can be made in those systems by a very low-energy signal because it is treated as information in the context of a complex host computational machinery. These agents exhibit a degree of multi-scale plasticity that enables informational input to make strong changes in the structure of the cognitive system itself. The positive flip side of this vulnerability is that it avails those kinds of minds with a long term version of free will: the ability through practice and repeated effort to change their own thinking patterns, responses to stimuli, and functional cognition. This continuum is not meant to be a linear *scala naturae* that aligns with any kind of "direction" of evolutionary progress—evolution is free to move in any direction in this option space of cognitive capacity; instead, this scheme provides a way to formalize (for a pragmatic, engineering approach) the major transitions in cognitive capacity that can be exploited for increased insight and control. The goal of the scientist is to find the optimal position for a given system. Too far to the right, and one ends up attributing hopes and dreams to thermostats or simple AIs in a way that does not advance prediction and control. Too far to the left, and one loses the benefits of top-down control in favor of intractable micromanagement. Note also that this forms a continuum with respect to how much knowledge one has to have about the system's details in order to manipulate its function: for systems in class A, one has to know a lot about their workings to modify them. For class B, one has to know how to read-write the setpoint information, but does not need to know anything about how the system will implement those goals. For class C, one doesn't have to know how the system modifies its goal encodings in light of experience, because the system does all of this on its own—one only has to provide rewards and punishments. Images by Jeremy Guay of Peregrine Creative.

minds" *via* the efforts of bioengineers, which is sure to disrupt categories and conclusions that have been formed in the context of today's natural biosphere.

What of consciousness? It is likely impossible to understand sentience without understanding cognition, and the emphasis of this paper is on testable, empirical impacts of ways to understand cognition in all of its guises. By enabling the definition, detection, and comparison of cognition and intelligence, in diverse substrates beyond standard animals, we can enhance the range of embodiments in which sentience may result. In order to move the field forward *via* empirical progress, the focus of most of the discussion below is on ways to think about cognitive function, not on phenomenal or access consciousness [in the sense of the "Hard Problem" (Chalmers, 2013)]. However, I return to this issue at the end, discussing TAME's view of sentience as fundamentally tied to goal-directed activity, only some aspects of which can be studied *via* third person approaches.

The main goal is to help advance and delineate an exciting emerging field at the intersection of biology, philosophy, and the information sciences. By proposing a new framework and

examining it in a broad context of now physically realizable (not merely logically possible) living structures, it may be possible to bring conceptual, philosophical thought up to date with recent advances in science and technology. At stake are current knowledge gaps in evolutionary, developmental, and cell biology, a new roadmap for regenerative medicine, lessons that could be ported to artificial intelligence and robotics, and broader implications for ethics.

COGNITION: CHANGING THE SUBJECT

Even advanced animals are really collective intelligences (Couzin, 2007, 2009; Valentini et al., 2018), exploiting still poorly-understood scaling and binding features of metazoan architectures that share a continuum with looser swarms that have been termed "liquid brains" (Sole et al., 2019). Studies of "centralized control" focus on a brain, which is in effect a network of cells performing functions that many cell types, including bacteria, can do (Koshland, 1983). The embodied nature of cognition means that the minds of Selves are dependent on a highly plastic material substrate which changes not only on evolutionary time scales but also during the lifetime of the agent itself.

The central consequence of the composite nature of all intelligences is that the Self is subject to significant change in real-time (**Figure 3**). This means both slow maturation through experience (a kind of "software" change that doesn't disrupt traditional ways of thinking about agency), as well as radical changes of the material in which a given mind is implemented (Levin, 2020). The owner, or subject of memories, preferences, and in more advanced cases, credit and blame, is very malleable. At the same time, fascinating mechanisms somehow ensure the persistence of Self (such as complex memories) despite drastic alterations of substrate. For example, the massive remodeling of the caterpillar brain, followed by the morphogenesis of an entirely different brain suitable for the moth or beetle, does not wipe all the memories of the larva but somehow maps them onto behavioral capacities in the post-metamorphosis host, despite its entirely different body (Alloway, 1972; Tully et al., 1994; Sheiman and Tiras, 1996; Armstrong et al., 1998; Ray, 1999; Blackiston et al., 2008). Not only that, but memories can apparently persist following the complete regeneration of brains in some organisms (McConnell et al., 1959; Corning, 1966; Shomrat and Levin, 2013) such as planaria, in which prior knowledge and behavioral tendencies are somehow transferred onto a newly-constructed brain. Even in vertebrates, such as fish (Versteeg et al., 2021) and mammals (von der Ohe et al., 2006), brain size and structure can change repeatedly during their lifespan. This is crucial to understanding agency and intelligence at multiple scales and in unfamiliar embodiments because observations like this begin to break down the notion of Selves as monadic, immutable objects with a privileged scale. Becoming comfortable with biological cognitive agents that are malleable in terms of form and function (change radically during the lifetime of an individual) makes it easier to understand the origins and changes of cognition during evolution or as the result of bioengineering effort.

This little-studied intersection between regeneration/remodeling and cognition highlights the fascinating plasticity of the body, brain, and mind; traditional model systems in which cognition is mapped onto a stable, discrete, mature brain are insufficient to fully understand the relationship between the Self and its material substrate. Many scientists study the behavioral properties of caterpillars, and of butterflies, but the transition zone in-between, from the perspective of philosophy of mind and cognitive science, provides an important opportunity to study the mind-body relationship by changing the body during the lifetime of the agent (not just during evolution). Note that continuity of being across drastic biological remodeling is not only relevant for unusual cases in the animal kingdom, but is a fundamental property of most life—even humans change from a collection of cells to a functional individual, *via* a gradual morphogenetic process that constructs an active Self in real time. This has not been addressed in biology, and likewise not yet in computer science, where machine learning approaches use static neural networks (there is not a formalism for altering artificial neural networks' architecture on the fly).

What are the invariants that enable a Self to persist (and be recognizable by third-person investigations) despite such change? Memory is a good candidate (Shoemaker, 1959; Ameriks, 1976; **Figure 3**). However, at least certain kinds of memories can be transferred between individuals, by transplants of brain tissue or molecular engrams (Pietsch and Schneider, 1969; McConnell and Shelby, 1970; Bisping et al., 1971; Chen et al., 2014; Bedecarrats et al., 2018; Abraham et al., 2019). Importantly, the movement of memories across individual animals is only a special case of the movement of memory in biological tissue in general. Even when housed in the same "body," memories must move between tissues—for example, in a trained planarian's tail fragment re-imprinting its learned information onto the newly regenerated brain, or the movement of memories onto new brain tissue during metamorphosis. In addition to the spatial movement and re-mapping of memories onto new substrates, there is also a temporal component, as each memory is really an instance of communication between past and future Selves. The plasticity of biological bodies, made of cells that die, are born, and significantly rearrange their tissue architecture, suggests that the understanding of cognition is fundamentally a problem of collective intelligence: to understand how stable cognitive structures can persist and map onto swarm *dynamics*, with preferences and stressors that scale from those of their components.

This is applicable even to such a "stable" form as the human brain, which is often spoken of as a single Subject of experience and thought. First, the gulf between planarian regeneration/insect metamorphosis and human brains is going to be bridged by emerging therapeutics. It is inevitable that stem cell therapies for degenerative brain diseases (Forraz et al., 2013; Rosser and Svendsen, 2014; Tanna and Sachan, 2014) will confront us with humans whose brains are partially replaced by the naïve progeny of cells that were not present during the formation of memories and personality traits in the patient. Even prior to these advances, it was clear that phenomena such as dissociative identity disorder (Miller and Triggiano, 1992), communication

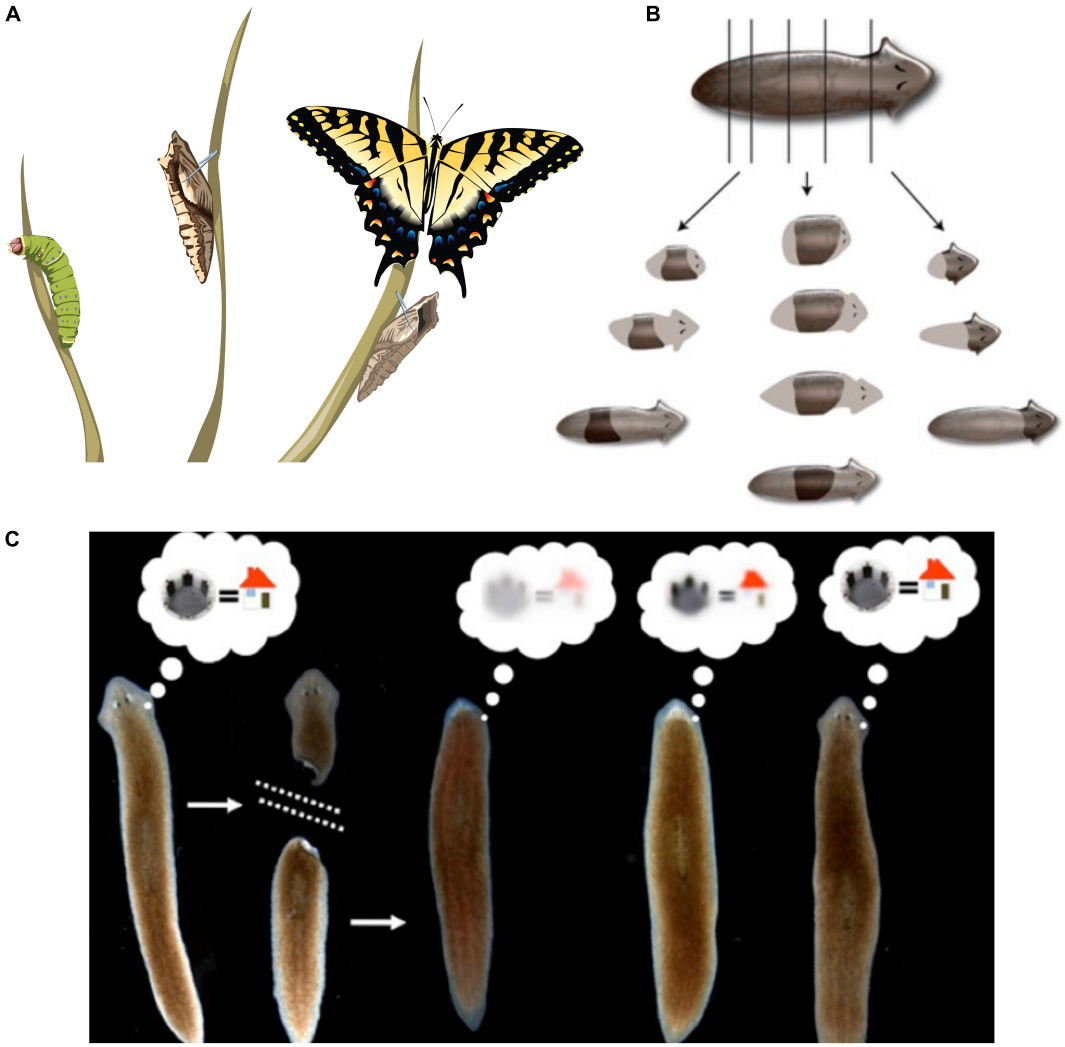


FIGURE 3 | Cognitive Selves can change in real-time. **(A)** Caterpillars metamorphose into butterflies, going through a process in which their body, brain, and cognitive systems are drastically remodeled during the lifetime of a single agent. Importantly, memories remain and persist through this process (Blackiston et al., 2015). **(B)** Planaria cut into pieces regenerate, with each piece re-growing and remodeling precisely what is needed to form an entire animal. **(C)** Planarians derived from tail fragments of trained worms still retain original information, illustrating the ability of memories to move across tissues and be reimprinted on newly-developing brains (Corning, 1966, 1967; Shomrat and Levin, 2013). Images by Jeremy Guay of Peregrine Creative.

with non-verbal brain hemispheres in commissurotomy patients (Nagel, 1971; Montgomery, 2003), conjoined twins with fused brains (Gazzaniga, 1970; Barilan, 2003), etc., place human cognition onto a continuous spectrum with respect to the plasticity of integrated Selves that reside within a particular biological tissue implementation.

Importantly, animal model systems are now providing the ability to harness that plasticity for functional investigations of the body-mind relationship. For example, it is now easy to radically modify bodies in a time-scale that is much faster than evolutionary change, to study the inherent plasticity of minds without eons of selection to shape them to fit specific body architectures. When tadpoles are created to have eyes on their tails, instead of their heads, they are still readily able to perform visual learning tasks (Blackiston and Levin, 2013; Blackiston et al., 2017). Planaria can readily be made with two (or more) brains in the same body (Morgan, 1904; Oviedo et al., 2010), and human patients are now routinely augmented with novel inputs [such as sensory substitution (Bach-y-Rita et al., 1969; Bach-y-Rita, 1981; Danilov and Tyler, 2005; Ptito et al., 2005)] or novel effectors, such as instrumentized interfaces allowing thought to control engineered devices such as wheelchairs in addition to the default muscle-driven peripherals of their own bodies (Green and Kalaska, 2011; Chamola et al., 2020; Belwafi et al., 2021). The central phenomenon here is plasticity: minds are not tightly bound to one specific underlying architecture (as most of our software is today), but readily mold to changes of genomic defaults. The logical extension of this progress is a focus on self-modifying living beings and the creation of new agents in which the mind:body system is

simplified by entirely replacing one side of the equation with an engineered construct. The benefit would be that at least one half of the system is now well-understood.

For example, in hybrots, animal brains are functionally connected to robotics instead of their normal body (Reger et al., 2000; Potter et al., 2003; Tsuda et al., 2009; Ando and Kanzaki, 2020). It doesn't even have to be an entire brain—a plate of neurons can learn to fly a flight simulator, and it lives in a new virtual world (DeMarse and Dockendorf, 2005; Manicka and Harvey, 2008; Beer, 2014), as seen from the development of closed-loop neurobiological platforms (Demarse et al., 2001; Potter et al., 2005; Bakkum et al., 2007b; Chao et al., 2008; Rolston et al., 2009a,b). These kinds of results are reminiscent of Philosophy 101's "brain in a vat" experiment (Harman, 1973). Brains adjust to driving robots and other devices as easily as they adjust to controlling a typical, or highly altered, living body because minds are somehow adapted and prepared to deal with body alterations—throughout development, metamorphosis and regeneration, and evolutionary change.

The massive plasticity of bodies, brains, and minds means that a mature cognitive science cannot just concern itself with understanding standard "model animals" as they exist right now. The typical "subject," such as a rat or fruit fly, which remains constant during the course of one's studies and is conveniently abstracted as a singular Self or intelligence, obscures the bigger picture. The future of this field must expand to frameworks that can handle all of the possible minds across an immense option space of bodies. Advances in bioengineering and artificial intelligence suggest that we or our descendants will be living in a world in which Darwin's "endless forms most beautiful" (this Earth's $N = 1$ ecosystem outputs) are just a tiny sample of the true variety of possible beings. Biobots, hybrots, cyborgs, synthetic and chimeric animals, genetically and cellularly bioengineered living forms, humans instrumentized to knowledge platforms, devices, and each other—these technologies are going to generate beings whose body architectures are nothing like our familiar phylogeny. They will be a functional mix of evolved and designed components; at all levels, smart materials, software-level systems, and living tissue will be integrated into novel beings which function in their own exotic Umwelt. Importantly, the information that is used to specify such beings' form and function is no longer only genetic—it is truly "epigenetic" because it comes not only from the creature's own genome but also from human and non-human agents' minds (and eventually, robotic machine-learning-driven platforms) that use cell-level bioengineering to generate novel bodies from genetically wild-type cells. In these cases, the genetics are no guide to the outcome (which highlights some of the profound reasons that genetics is hard to use to truly predict cognitive form and function even in traditional living species).

Now is the time to begin to develop ways of thinking about truly novel bodies and minds, because the technology is advancing more rapidly than philosophical progress. Many of the standard philosophical puzzles concerning brain hemisphere transplants, moving memories, replacing body/brain parts, etc. are now eminently doable in practice, while the theory of how to interpret the results lags. We now have the opportunity to begin to develop conceptual approaches to (1) understand beings without convenient evolutionary back-stories as explanations for their cognitive capacities (whose minds are created *de novo*, and not shaped by long selection pressures toward specific capabilities), and (2) develop ways to analyze novel Selves that are not amenable to simple comparisons with related beings, not informed by their phylogenetic position relative to known standard species, and not predictable from an analysis of their genetics. The implications range across insights into evolutionary developmental biology, advancing bioengineering and artificial life research, new roadmaps for regenerative medicine, ability to recognize exobiological life, and the development of ethics for relating to novel beings whose composition offers no familiar phylogenetic touchstone. Thus, here I propose the beginnings of a framework designed to drive empirical research and conceptual/philosophical analysis that will be broadly applicable to minds regardless of their origin story or internal architecture.

TECHNOLOGICAL APPROACH TO MIND EVERYWHERE: A PROPOSAL FOR A FRAMEWORK

The Technological Approach to Mind Everywhere (TAME) framework seeks to establish a way to recognize, study, and compare truly diverse intelligences in the space of possible agents. The goal of this project is to identify deep invariants between cognitive systems of very different types of agents, and abstract away from inessential features such as composition or origin, which were sufficient heuristics with which to recognize agency in prior decades but will surely be insufficient in the future (Bongard and Levin, 2021). To flesh out this approach, I first make explicit some of its philosophical foundations, and then discuss specific conceptual tools that have been developed to begin the task of understanding embodied cognition in the space of mind-as-it-can-be (a sister concept to Langton's motto for the artificial life community—"life as it can be") (Langton, 1995).

Philosophical Foundations of an Approach to Diverse Intelligences

One key pillar of this research program is the commitment to gradualism with respect to almost all important cognition-related properties: advanced minds are in important ways generated in a continuous manner from much more humble proto-cognitive systems. On this view, it is hopeless to look for a clear bright line that demarcates "true" cognition (such as that of humans, great apes, etc.) from metaphorical "as if cognition" or "just physics." Taking evolutionary biology seriously means that there is a continuous series of forms that connect any cognitive system with much more humble ones. While phylogenetic history already refutes views of a magical arrival of "true cognition" in one generation, from parents that didn't have it (instead stretching the process of cognitive expansion over long time scales and slow modification), recent advances in biotechnology make this completely implausible. For any putative difference between a creature that is proposed to have *true* preferences, memories, and

plans and one that supposedly has *none*, we can now construct in-between, hybrid forms which then make it impossible to say whether the resulting being is an Agent or not. Many pseudo-problems evaporate when a binary view of cognition is dissolved by an appreciation of the plasticity and interoperability of living material at all scales of organization. A definitive discussion of the engineering of preferences and goal-directedness, in terms of hierarchy requirements and upper-directedness, is given in McShea (2013, 2016).

For example, one view is that only biological, evolved forms have intrinsic motivation, while software AI agents are only faking it *via* functional performance [but don't actually *care* (Oudeyer and Kaplan, 2007, 2013; Lyon and Kuchling, 2021)]. But which biological systems *really* care—fish? Single cells? Do mitochondria (which used to be independent organisms) have true preferences about their own or their host cells' physiological states? The lack of consensus on this question in classical (natural) biological systems, and the absence of convincing criteria that can be used to sort all possible agents to one or the other side of a sharp line, highlight the futility of truly binary categories. Moreover, we can now readily construct hybrid systems that consist of any percentage of robotics tightly coupled to on-board living cells and tissues, which function together as one integrated being. How many living cells does a robot need to contain before the living system's "true" cognition bleeds over into the whole? On the continuum between human brains (with electrodes and a machine learning converter chip) that drive assistive devices (e.g., 95% human, 5% robotics), and robots with on-board cultured human brain cells instrumentized to assist with performance (5% human, 95% robotics), where can one draw the line—given that any desired percent combination is possible to make? No quantitative answer is sufficient to push a system "over the line" because there is no such line (at least, no convincing line has been proposed). Interesting aspects of agency or cognition are rarely if ever Boolean values.

Instead of a binary dichotomy, which leads to impassable philosophical roadblocks, we envision a continuum of advancement and diversity in information-processing capacity. Progressively more complex capabilities [such as unlimited associative learning, counterfactual modeling, symbol manipulation, etc., (Ginsburg and Jablonka, 2021)] ramp up, but are nevertheless part of a continuous process that is not devoid of proto-cognitive capacity before complex brains appear. Specifically, while major differences in cognitive function of course exist among diverse intelligences, transitions between them have not been shown to be binary or rapid relative to the timescale of individual agents. There is no plausible reason to think that evolution produces parents that don't have "true cognition" but give rise to offspring that suddenly do, or that development starts with an embryo that has no "true preferences" and sharply transitions into an animal that does, etc. Moreover, bioengineering and chimerization can produce a smooth series of transitional forms between any two forms that are proposed to have, or not have, any cognitive property. Thus, agents gradually shift (during their lifetime, as result of development, metamorphosis, or interactions with other agents, or during evolutionary timescales) between great transitions in cognitive capacity, expressing and experiencing intermediate states of cognitive capacity that must be recognized by empirical approaches to study them.

A focus on the plasticity of the embodiments of mind strongly suggests this kind of gradualist view, which has been expounded in the context of evolutionary forces controlling individuality (Godfrey-Smith, 2009; Queller and Strassmann, 2009). Here the additional focus is on events taking place within the lifetime of individuals and driven by information and control dynamics. The TAME framework pushes experimenters to ask "how much" and "what kind of" cognition any given system might manifest if we interacted with it in the right way, at the right scale of observation. And of course, the degree of cognition is not a single parameter that gives rise to a *scala naturae* but a shorthand for the shape and size of its cognitive capacities in a rich space (discussed below).

The second pillar of TAME is that there is no privileged material substrate for Selves. Alongside familiar materials such as brains made of neurons, the field of basal cognition (Nicolis et al., 2011; Reid et al., 2012, 2013; Beekman and Latty, 2015; Baluška and Levin, 2016; Boussard et al., 2019; Dexter et al., 2019; Gershman et al., 2021; Levin et al., 2021; Lyon et al., 2021) has been identifying novel kinds of intelligences in single cells, plants, animal tissues, and swarms. The fields of active matter, intelligent materials, swarm robotics, machine learning, and someday, exobiology, suggest that we cannot rely on a familiar signature of "big vertebrate brain" as a necessary condition for mind. Molecular phylogeny shows that the specific components of brains pre-date the evolution of neurons *per se*, and life has been solving problems long before brains came onto the scene (Buznikov et al., 2005; Levin et al., 2006; Jekely et al., 2015; Liebeskind et al., 2015; Moran et al., 2015). Powerful unification and generalization of concepts from cognitive science and other fields can be achieved if we develop tools to characterize and relate to a wide diversity of minds in unconventional material implementations (Damasio, 2010; Damasio and Carvalho, 2013; Cook et al., 2014; Ford, 2017; Man and Damasio, 2019; Baluska et al., 2021; Reber and Baluska, 2021).

Closely related to that is the de-throning of natural evolution as the only acceptable origin story for a true Agent [many have proposed a distinction between evolved living forms vs. the somehow inadequate machines which were merely designed by man (Bongard and Levin, 2021)]. First, synthetic evolutionary processes are now being used in the lab to create "machines" and modify life (Kriegman et al., 2020a; Blackiston et al., 2021). Second, the whole process of evolution, basically a hill-climbing search algorithm, results in a set of frozen accidents and meandering selection among random tweaks to the micro-level hardware of cells, with impossible to predict large-scale consequences for the emergent system level structure and function. If this short-sighted process, constrained by many forces that have nothing to do with favoring complex cognition, can give rise to true minds, then so can a rational engineering approach. There is nothing magical about evolution (driven by randomizing processes) as a forge for cognition; surely we can eventually do at least as well, and likely much

better, using rational construction principles and an even wider range of materials.

The third foundational aspect of TAME is that the correct answer to how much agency a system has cannot be settled by philosophy—it is an empirical question. The goal is to produce a framework that drives experimental research programs, not only philosophical debate about what should or should not be possible as a matter of definition. To this end, the productive way to think about this a variant of Dennett's Intentional Stance (Dennett, 1987; Mar et al., 2007), which frames properties such as cognition as observer-dependent, empirically testable, and defined by how much benefit their recognition offers to science (**Figure 2**). Thus, the correct level of agency with which to treat any system must be determined by experiments that reveal which kind of model and strategy provides the most efficient predictive and control capability over the system. In this engineering (understand, modify, build)-centered view, the optimal position of a system on the spectrum of agency is determined empirically, based on which kind of model affords the most efficient way of prediction and control. Such estimates are, by their empirical nature, subject to revision by future experimental data and conceptual frameworks, and are observer-dependent (not absolute).

A standard methodology in science is to avoid attributing agency to a given system unless absolutely necessary. The mainstream view (e.g., Morgan's Canon) is that it's too easy to fall into a trap of "anthropomorphizing" systems with only apparent cognitive powers, when one should only be looking for models focused on mechanistic, lower levels of description that eschew any kind of teleology or mental capacity (Morgan, 1903; Epstein, 1984). However, analysis shows that this view provides no useful parsimony (Cartmill, 2017). The rich history of debates on reductionism and mechanism needs to be complemented with an empirical, engineering approach that is not inappropriately slanted in one direction on this continuum. Teleophobia leads to Type 2 errors with respect to attribution of cognition that carry a huge opportunity cost for not only practical outcomes like regenerative medicine (Pezzulo and Levin, 2015) and engineering, but also ethics. Humans (and many other animals) readily attribute agency to systems in their environment; scientists should be comfortable with testing out a theory of mind regarding various complex systems for the exact same reason—it can often greatly enhance prediction and control, by recognizing the true features of the systems with which we interact. This perspective implies that there is no such thing as "anthropomorphizing" because human beings have no unique essential property which can be inappropriately attributed to agents that have *none* of it. Aside from the very rare trivial cases (misattributing human-*level* cognition to simpler systems), we must be careful to avoid the pervasive, implicit remnants of a human-centered pre-scientific worldview in which modern, standard humans are assumed to have some sort of irreducible quality that cannot be present in degrees in slightly (or greatly) different physical implementations (from early hominids to cyborgs etc.). Instead, we should seek ways to naturalize human capacities as elaborations of more fundamental principles that are widely present in complex systems, in very different types and degrees, and to identify the *correct* level for any given system. Of course, this is just one stance, emphasizing experimental, not philosophical, approaches that avoid defining impassable absolute differences that are not explainable by any known binary transition in body structure or function. Others can certainly drive empirical work focused specifically on what kind of human-level capacities do and do not exist in detectable quantity in other agents.

Avoiding philosophical wrangling over privileged levels of explanation (Ellis, 2008; Ellis et al., 2012; Noble, 2012), TAME takes an empirical approach to attributing agency, which increases the toolkit of ways to relate to complex systems, and also works to reduce profligate attributions of mental qualities. We do not say that a thermos knows whether to keep something hot or cold, because no model of thermos cognition does better than basic thermodynamics to explain its behavior or build better thermoses. At the same time, we know we cannot simply use Newton's laws to predict the motion of a (living) mouse at the top of a hill, requiring us to construct models of navigation and goal-directed activity for the controller of the mouse's behavior over time (Jennings, 1906).

Under-estimating the capacity of a system for plasticity, learning, having preferences, representation, and intelligent problem-solving greatly reduces the toolkit of techniques we can use to understand and control its behavior. Consider the task of getting a pigeon to correctly distinguish videos of dance vs. those of martial arts. If one approaches the system bottom-up, one has to implement ways to interface to individual neurons in the animal's brain to read the visual input, distinguish the videos correctly, and then control other neurons to force the behavior of walking up to a button and pressing it. This may someday be possible, but not in our lifetimes. In contrast, one can simply train the pigeon (Qadri and Cook, 2017). Humanity has been training animals for millennia, without knowing anything about what is in their heads or how brains work. This highly efficient trick works because we correctly identified them as learning agents, which allows us to offload a lot of the computational complexity of any task onto the living system itself, without micromanaging its components.

What other systems might this remarkably powerful strategy apply to? For example, gene regulatory networks (GRNs) are a paradigmatic example of "genetic mechanism," often assumed to be tractable only by hardware (requiring gene therapy approaches to alter promoter sequences that control network connectivity, or adding/removing gene nodes). However, being open to the possibility that GRNs might actually be on a different place on this continuum suggests an experiment in which they are trained for new behaviors with specific combinations of stimuli (experiences). Indeed, recent analyses of biological GRN models reveal that they exhibit associative and several other kinds of learning capacity, as well as pattern completion and generalization (Watson et al., 2010, 2014; Szilagyi et al., 2020; Biswas et al., 2021). This is an example in which an empirical approach to the correct level of agency for even simple systems not usually thought of as cognitive suggests new hypotheses which in turn open a path to new practical applications (biomedical strategies using associative regimes of

drug pulsing to exploit memory and address pharmacoresistance by abrogating habituation, etc.).

We next consider specific aspects of the framework, before diving into specific examples in which it drives novel empirical work.

Specific Conceptual Components of the Technological Approach to Mind Everywhere Framework

A useful framework in this emerging field should not only serve as a lens with which to view data and concepts (Manicka and Levin, 2019b), but also should drive research in several ways. It needs to first specify definitions for key terms such as a Self. These are not meant to be exclusively correct—the definitions can co-exist with others, but should identify a claim as to what is an essential invariant for Selves (and what other aspects can diverge), and how it intersects with experiment. The fundamental symmetry unifying all possible Selves should also facilitate direct comparison or even classification of truly diverse intelligences, sketching the markers of Selfhood and the topology of the option space within which possible agents exist. The framework should also help scientists derive testable claims about how borders of a given Self are determined, and how it interacts with the outside world (and other agents). Finally, the framework should provide actionable, semi-quantitative definitions that have strong implications and constrain theories about how Selves arise and change. All of this must facilitate experimental approaches to determine the empirical utility of this approach.

The TAME framework takes the following as the basic hallmarks of being a Self: the ability to pursue goals, to own compound (e.g., associative) memories, and to serve as the locus for credit assignment (be rewarded or punished), where all of these are at a scale larger than possible for any of its components alone. Given the gradualist nature of the framework, the key question for any agent is "how well," "how much," and "what kind" of capacity it has for each of those key aspects, which in turn allows agents to be directly compared in an option space. TAME emphasizes defining a higher scale at which the (possibly competent) activity of component parts gives rise to an emergent system. Like a valid mathematical theorem which has a unique structure and existence over and above any of its individual statements, a Self can own, for example, associative memories (that bind into new mental content experiences that occurred separately to its individual parts), be the subject of reward or punishment for complex states (as a consequence of highly diverse actions that its parts have taken), and be stressed by states of affairs (deviations from goals or setpoints) that are not definable at the level of any of its parts (which of course may have their own distinct types of stresses and goals). These are practical aspects that suggest ways to recognize, create, and modify Selves.

Selves can be classified and compared with respect to the scale of goals they can pursue [**Figure 4**, described in detail in Levin (2019)]. In this context, the goal-directed perspective adopted here builds on the work of Rosenblueth et al. (1943); Nagel (1979); and Mayr (1992), emphasizing plasticity (ability to reach a goal state from different starting points) and persistence (capacity to reach a goal (Schlosser, 1998) state despite perturbations).

The ability of a system to exert energy to work toward a state of affairs, overcoming obstacles (to the degree that its sophistication allows) to achieve a particular set of substates is very useful for defining Selves because it grounds the question in well-established control theory and cybernetics (i.e., systems "trying to do things" is no longer magical but is well-established in engineering), and provides a natural way of discovering, defining, and altering the preferences of a system. A common objection is: "surely we can't say that thermostats have goals and preferences?" The TAME framework holds that whatever true goals and preferences are, there must exist primitive, minimal versions from which they evolved and these are, in an important sense, substrate- and scale-independent; simple homeostatic circuits are an ideal candidate for the "hydrogen atom" of goal-directed activity (Rosenblueth et al., 1943; Turner, 2019). A key tool for thinking about these problems is to ask what a truly minimal example of any cognitive capacity would be like, and to think about transitional forms that can be created just below that. It is logically inevitable that if one follows a complex cognitive capacity backward through phylogeny, one eventually reaches precursor versions of that capacity that naturally suggest the (misguided) question "is that *really* cognitive, or just physics?" Indeed, a kind of minimal goal-directedness permeates all of physics (Feynman, 1942; Georgiev and Georgiev, 2002; Ogborn et al., 2006; Kaila and Annila, 2008; Ramstead et al., 2019; Kuchling et al., 2020a), supporting a continuous climb of the scale and sophistication of goals.

Pursuit of goals is central to composite agency and the "many to one" problem because it requires distinct mechanisms (for measurement of states, storing setpoints, and driving activity to minimize the delta between the former and the latter) to be bound together into a functional unit that is greater than its parts. To co-opt a great quote (Dobzhansky, 1973), nothing in biology makes sense except in light of teleonomy (Pittendrigh, 1958; Nagel, 1979; Mayr, 1992; Schlosser, 1998; Noble, 2010, 2011; Auletta, 2011; Ellis et al., 2012). The degree to which a system can evaluate possible consequences of various actions, in pursuit of those goal states, can vary widely, but is essential to its survival. The expenditure of energy *in ways that effectively reach specific states despite uncertainty, limitations of capability, and meddling from outside forces* is proposed as a central unifying invariant for all Selves—a basis for the space of possible agents. This view suggests a semi-quantitative multi-axis option space that enables direct comparison of diverse intelligences of all sorts of material implementation and origins (Levin, 2019, 2020). Specifically (**Figure 4**), a "space-time" diagram can be created where the spatio-temporal *scale* of any agent's goals delineates that Self and its cognitive boundaries.

Note that the distances on **Figure 4D** represent not first-order capacities such as sensory perception (how far away can it sense), but second-order capacities of the size of goals (humble metabolic hunger-satiety loops or grandiose planetary-scale engineering ambitions) which a given cognitive system is capable of representing and working toward. At any given time, an Agent is represented by a single shape in this space,

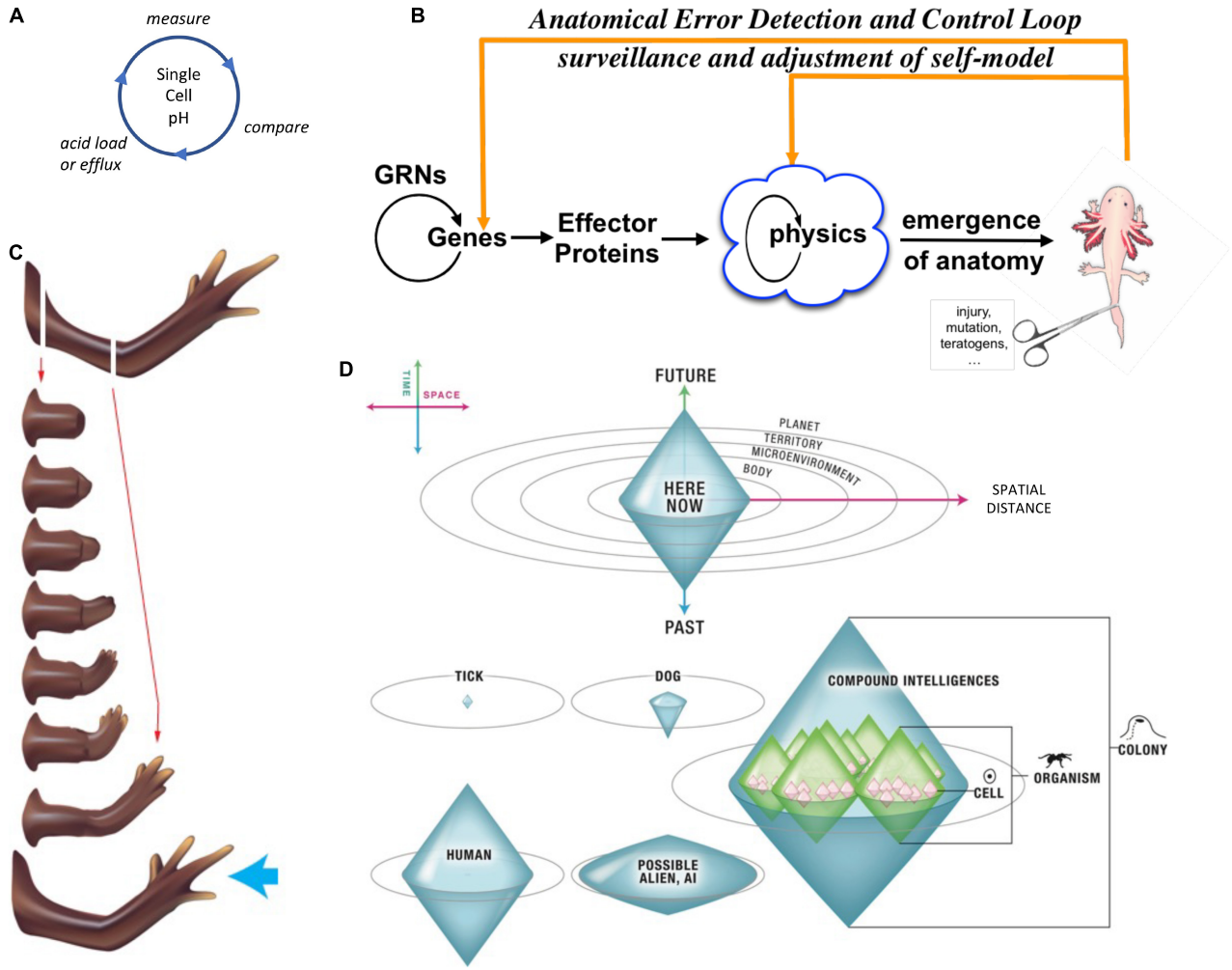


FIGURE 4 | Unconventional goal-directed agents and the scaling of the cognitive Self. **(A)** The minimal component of agency is homeostasis, for example the ability of a cell to execute the Test-Operate-Exit (Pezzulo and Levin, 2016) loop: a cycle of comparison with setpoint and adjustment *via* effectors, which allows it to remain in a particular region of state space. **(B)** This same capacity is scaled up by cellular networks into anatomical homeostasis: morphogenesis is not simply a feedforward emergent process but rather the ability of living systems to adjust and remodel to specific target morphologies. This requires feedback loops at the transcriptional and biophysical levels, which rely on stored information (e.g., bioelectrical pattern memories) against which to minimize error. **(C)** This is what underlies complex regeneration such as salamander limbs, which can be cut at any position and result in just the right amount and type of regenerative growth that stops when a correct limb is achieved. Such homeostatic systems are examples of simple goal-directed agents. **(D)** A focus on the size or scale of goals any given system can pursue allows plotting very diverse intelligences on the same graph, regardless of their origin or composition (Levin, 2019). The scale of their goal-directed activity is estimated (collapsed onto one axis of space and one of time, as in Relativity diagrams). Importantly, this way of visualizing the sophistication of agency is a schematic of goal space—it is not meant to represent the spatial extent of sensing or effector range, but rather the scale of events about which they care and the boundary of states that they can possibly represent or work to change. This defines a kind of cognitive light cone (a boundary to any agent's area of concern); the largest area represents the "now," with fading efficacy both backward (accessing past events with decreasing reliability) and forward (limited prediction accuracy for future events). Agents are compound entities, composed of (and comprising) other sub- or super-agents each of which has their own cognitive boundary of various sizes. Images by Jeremy Guay of Peregrine Creative.

corresponding to the size and complexity of their possible goal domain. However, genomes (or engineering design specs) map to an ensemble of such shapes in this space because the borders between Self and world, and the scope of goals an agent's cognitive apparatus can handle, can all shift during the lifetime of some agents—"in software" (another "great transition" marker). All regions in this space can potentially define some possible agent. Of course, additional subdivisions (dimensions) can easily be added, such as the Unlimited Associative Learning marker (Birch et al., 2020) or aspects of Active Inference (Friston and Ao, 2012; Friston et al., 2015b; Calvo and Friston, 2017; Peters et al., 2017).

Some agents, like microbes, have minimal memory (Vladimirov and Sourjik, 2009; Lan and Tu, 2016) and can concern themselves only with a very short time horizon and spatial radius—e.g., follow local gradients. Some agents, e.g., a rat have more memory and some forward planning ability (Hadj-Chikh et al., 1996; Raby and Clayton, 2009;

Smith and Litchfield, 2010), but are still precluded from, for example, effectively caring about what will happen 2 months hence, in an adjacent town. Some, like human beings, can devote their lives to causes of enormous scale (future state of the planet, humanity, etc.). Akin to Special Relativity, this formalization makes explicit that class of capacities (in terms of representation of classes of goals) that are forever inaccessible to a given agent (demarcating the edge of the "light cone" of its cognition).

In general, larger selves (1) are capable of working toward states of affairs that occur farther into the future (perhaps outlasting the lifetime of the agent itself—an important great transition, in the sense of West et al. (2015), along the cognitive continuum); (2) deploy memories further back in time (their actions become less "mechanism" and more *decision-making* (Balazsi et al., 2011) because they are linked to a network of functional causes and information with larger diameter); and (3) they expend effort to manage sensing/effector activity in larger spaces [from subcellular networks to the extended mind (Clark and Chalmers, 1998; Turner, 2000; Timsit and Gregoire, 2021)]. Overall, increases of agency are driven by mechanisms that scale up stress (**Box 1**)—the scope of states that an agent can possibly be stressed about (in the sense of pressure to take corrective action). In this framework, stress (as a system-level response to distance from setpoint states), preferences, motivation, and the ability to functionally care about what happens are tightly linked. Homeostasis, necessary for life, evolves into allostasis (McEwen, 1998; Schulkin and Sterling, 2019) as new architectures allow tight, local homeostatic loops to be scaled up to measure, cause, and remember larger and more complex states of affairs (Di Paulo, 2000; Camley, 2018).

Additional implications of this view are that Selves: are malleable (the borders and scale of any Self can change over time); can be created by design or by evolution; and are multi-scale entities that consist of other, smaller Selves (and conversely, scale up to make larger Selves). Indeed they are a patchwork of agents [akin to Theophile Bordeu's "many little lives" (Haigh, 1976; Wolfe, 2008)] that overlap with each other, and compete, communicate, and cooperate both horizontally (at their own level of organization) and vertically [with their component subunits and the super-Selves of which they are a part (Sims, 2020)].

Another important invariant for comparing diverse intelligences is that they are all solving problems, in some space (**Figure 5**). It is proposed that the traditional problem-solving behavior we see in standard animals in 3D space is just a variant of evolutionarily more ancient capacity to solve problems in metabolic, physiological, transcriptional, and morphogenetic spaces (as one possible sequential timeline along which evolution pivoted some of the same strategies to solve problems in new spaces). For example, when planaria are exposed to barium, a non-specific potassium channel blocker, their heads explode. Remarkably, they soon regenerate heads that are completely insensitive to barium (Emmons-Bell et al., 2019). Transcriptomic analysis revealed that relatively few genes out of the entire genome were regulated to enable the cells to resolve this physiological stressor using transcriptional effectors to change how ions and neurotransmitters are handled by the cells. Barium is not something planaria ever encounter ecologically (so there should not be innate evolved responses to barium exposure), and cells don't turn over fast enough for a selection process (e.g., with bacterial persisters after antibiotic exposure). The task of determining which genes, out of the entire genome, can be transcriptionally regulated to return to an appropriate physiological regime is an example of an unconventional intelligence navigating a large-dimensional space to solve problems in real-time (Voskoboynik et al., 2007; Elgart et al., 2015; Soen et al., 2015; Schreier et al., 2017). Also interesting is that the actions taken in transcriptional space (a set of mRNA states) map onto a path in physiological state (the *ability* to perform many needed functions despite abrogated K^+ channel activity, not just a single state).

The common feature in all such instances is that the agent must navigate its space(s), preferentially occupying adaptive regions despite perturbations from the outside world (and from internal events) that tend to pull it into novel regions. Agents (and their sub- and super-agents) construct internal models of

BOX 1 | Stress as the glue of agency.

Tell me what you are stressed about and I will know a lot about your cognitive sophistication. Local glucose concentration? Limb too short? Rival is encroaching on your territory? Your limited lifespan? Global disparities in quality of life on Earth? The scope of states that an agent can possibly be stressed by, in effect, defines their degree of cognitive capacity. Stress is a systemic response to a difference between current state and a desired setpoint; it is an essential component to scaling of Selves because it enables different modules (which sense and act on things at different scales and in distributed locations) to be bound together in one global homeostatic loop (toward a larger purpose). Systemic stress occurs when one sub-agent is not satisfied about its local conditions, and propagates its unhappiness outward as hard-to-ignore signals. In this process, stress pathways serve the same function as hidden layers in a network, enabling the system to be more adaptive by connecting diverse modular inputs and outputs to the same basic stress minimization loop. Such networks scale stress, but stress is also what helps the network scale up its agency—a bidirectional positive feedback loop.

The key is that this stress signal is unpleasant to the other sub-agents, closely mimicking their own stress machinery (genetic conservation: my internal stress molecule is the same as your stress molecule, which contributes to the same "wiping of ownership" that is implemented by gap junctional connections). By propagating unhappiness in this way (in effect, turning up the global system "energy" which facilitates tendency for moving in various spaces), this process recruits distant sub-agents to act, to reduce their own perception of stress. For example, if an organ primordium is in the wrong location and needs to move, the surrounding cells are more willing to get out of the way if by doing so they reduce the amount of stress signal they receive. It may be a process akin to run-and-tumble for bacteria, with stress as the indicator of when to move and when to stop moving, in physiological, transcriptional, or morphogenetic space. Another example is compensatory hypertrophy, in which damage in one organ induces other cells to take up its workload, growing or taking on new functions if need be (Tamori and Deng, 2014; Fontes et al., 2020). In this way, stress causes other agents to work toward the same goal, serving as an influence that binds subunits across space into a coherent higher Self and resists the "struggle of the parts" (Heams, 2012). Interestingly, stress spreads not only horizontally in space (across cell fields) but also vertically, in time: effects of stress response is one of the things most easily transferred by transgenerational inheritance (Xue and Acar, 2018).

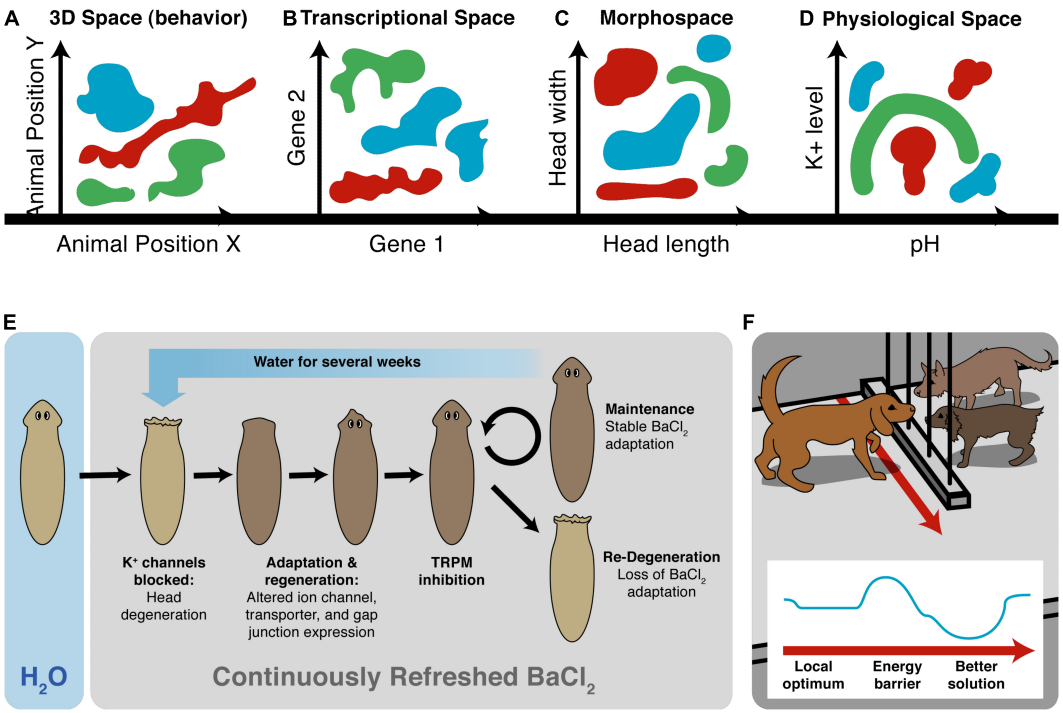


FIGURE 5 | Cognitive agents solve problems in diverse spaces. Intelligence is fundamentally about problem-solving, but this takes place not only in familiar 3D space as "behavior" (control of muscle effectors for movement) **(A)**, but also in other spaces in which cognitive systems try to navigate, in order to reach better regions. This includes the transcriptional space of gene expression **(B)** here schematized for two genes, anatomical morphospace **(C)** here schematized for two traits, and physiological space **(D)** here schematized for two parameters. An example **(E)** of problem-solving is planaria, which placed in barium (causing their heads to explode due to general blockade of potassium channels) regenerate new heads that are barium-insensitive (Emmons-Bell et al., 2019). They solve this entirely novel (not primed by evolutionary experience with barium) stressor by a very efficient traversal in transcriptional space to rapidly up/down regulate a very small number of genes that allows them to conduct their physiology despite the essential K^+ flux blockade. **(F)** The degree of intelligence of a system can be estimated by how effectively they navigate to optimal regions without being caught in a local maximum, illustrated as a dog which could achieve its goal on the other side of the fence, but this would require going around—temporarily getting further from its goal (a measurable degree of patience or foresight of any system in navigating its space, which can be visualized as a sort of energy barrier in the space, inset). Images by Jeremy Guay of Peregrine Creative.

their spaces (Beer, 2014, 2015; Beer and Williams, 2015; Hoffman et al., 2015; Fields et al., 2017; Hoffman, 2017; Prentner, 2019; Dietrich et al., 2020; Prakash et al., 2020), which may or may not match the view of their action space developed by their conspecifics, parasites, and scientists. Thus, the space one is navigating is in an important sense virtual (belonging to some Agent's self-model), is developed and often modified "on the fly" (in addition to that hardwired by the structure of the agent), and not only faces outward to infer a useful structure of its option space but also faces inward to map its own body and somatotopic properties (Bongard et al., 2006). The lower-level subsystems simplify the search space for the higher-level agent because their modular competency means that the higher-level system doesn't need to manage all the microstates [a strong kind of hierarchical modularity (Zhao et al., 2006; Lowell and Pollack, 2016)]. In turn, the higher-level system deforms the option space for the lower-level systems so that they do not need to be as clever, and can simply follow local energy gradients.

The degree of intelligence, or sophistication, of an agent in any space is roughly proportional to its ability to deploy memory and prediction (information processing) in order to avoid local maxima. Intelligence involves being able to temporarily move away from a simple vector toward one's goals in a way that results in bigger improvements down the line; the agent's internal complexity has to facilitate some degree of complexity (akin to hidden layers in an artificial neural network which introduce plasticity between stimulus and response) in the goal-directed activity that enables the buffering needed for patience and indirect paths to the goal. This buffering enables the flip side of homeostatic problem-driven (stress reduction) behavior by cells: the exploration of the space for novel opportunities (creativity) by the collective agent, and the ability to acquire more complex goals [in effect, beginning the climb to Maslow's hierarchy (Taormina and Gao, 2013)]. Of course it must be pointed out that this way of conceiving intelligence is one of many, and is proposed here as a way to enable the concept to be experimentally ported over to unfamiliar substrates, while capturing what is essential about it in a way that does not depend on arbitrary restrictions that will surely not survive advances in synthetic bioengineering, machine learning, and exobiology.

Another important aspect of intelligence that is space-agnostic is the capacity for generalization. For example, in the barium planaria example discussed above, it is possible that part of the problem-solving capacity is due to the cells' ability to generalize in

physiological space. Perhaps the cells recognize the physiological stresses induced by the novel barium stimulus as a member of the wider class of excitotoxicity induced by evolutionarily-familiar epileptic triggers, enabling them to deploy similar solutions (in terms of actions in transcriptional space). Such abilities to generalize have now been linked to measurement invariance (Frank, 2018), showing its ancient roots in the continuum of cognition.

Consistent with the above discussion, complex agents often consist of components that are themselves competent problem-solvers in their own (usually smaller, local) spaces. The relationship between wholes and their parts can be as follows. An agent is an integrated holobiont to the extent that it distorts the option space, and the geodesics through it, for its subunits (perhaps akin to how matter and space affect each other in general relativity) to get closer to a high-level goal in its space. A similar scheme is seen in neuroscience, where top-down feedback helps lower layer neurons to choose a response to local features by informing them about more global features (Krotov, 2021).

At the level of the subunits, which know nothing of the higher problem space, this simply looks like they are minimizing free energy and passively doing the only thing they can do as physical systems: this is why if one zooms in far enough on any act of decision-making, all one ever sees is dumb mechanism and "just physics." The agential perspective (Godfrey-Smith, 2009) looks different at different scales of observation (and its degree is in the eye of a beholder who seeks to control and predict the system, which includes the Agent itself, and its various partitions). This view is closely aligned with that of "upper directedness" (McShea, 2012), in which the larger system directs its components' behavior by constraints and rewards for coarse-grained outcomes, not microstates (McShea, 2012).

Note that these different competing and cooperating partitions are not just diverse components of the body (cells, microbiome, etc.) but also future and past versions of the Self. For example, one way to achieve the goal of a healthier metabolism is to lock the refrigerator at night and put the keys somewhere that your midnight self, which has a shorter cognitive boundary (is willing to trade long-term health for satiety right now) and less patience, is too lazy to find. Changing the option space, energy barriers, and reward gradients for your future self is a useful strategy for reaching complex goals despite the shorter horizons of the other intelligences that constitute your affordances in action space.

The most effective collective intelligences operate by simultaneously distorting the space to make it easy for their subunits to do the right thing with no comprehension of the larger-scale goals, but themselves benefit from the competency of the subunits which can often get their local job done even if the space is not perfectly shaped (because they themselves are homeostatic agents in their own space). Thus, instances of communication and control between agents (at the same or different levels) are mappings between different spaces. This suggests that both evolution's, and engineers', hard work is to optimize the appropriate functional mapping toward robustness and adaptive function.

Next, we consider a practical example of the application of this framework to an unconventional example of cognition and flexible problem-solving: morphogenesis, which naturally leads to specific hypotheses of the origin of larger biological Selves (scaling) and its testable empirical (biomedical) predictions (Dukas, 1998). This is followed with an exploration of the implications of these concepts for evolution, and a few remarks on consciousness.

SOMATIC COGNITION: AN EXAMPLE OF UNCONVENTIONAL AGENCY IN DETAIL

"Again and again terms have been used which point not to physical but to psychical analogies. It was meant to be more than a poetical metaphor. . ."

– Spemann (1967)

An example of TAME applied to basal cognition in an unconventional substrate is that of morphogenesis, in which the mechanisms of cognitive binding between subunits are now partially known, and testable hypotheses about cognitive scaling can be formulated [explored in detail in Friston et al. (2015a) and Pezzulo and Levin (2015, 2016)]. It is uncontroversial that morphogenesis is the result of collective activity: individual cells work together to build very complex structures. Most modern biologists treat it as clockwork [with a few notable exceptions around the recent data on cell learning (di Primio et al., 2000; Brugger et al., 2002; Norman et al., 2013; Yang et al., 2014; Stockwell et al., 2015; Urrios et al., 2016; Tweedy and Insall, 2020; Tweedy et al., 2020)], preferring a purely feed-forward approach founded on the idea of complexity science and emergence. On this view, there is a privileged level of causation—that of biochemistry—and all of the outcomes are to be seen as the emergent consequences of highly parallel execution of local rules (a cellular automaton in every sense of the term). Of course, it should be noted that the forefathers of developmental biology, such as Spemann (1967), were already well-aware of the possible role of cognitive concepts in this arena and others have occasionally pointed out detailed homologies (Grossberg, 1978; Pezzulo and Levin, 2015). This becomes clearer when we step away from the typical examples seen in developmental biology textbooks and look at some phenomena that, despite the recent progress in molecular genetics, remain important knowledge gaps (**Figure 6**).

Goal-Directed Activity in Morphogenesis

Morphogenesis (broadly defined) is not only a process that produces the same robust outcome from the same starting condition (development from a fertilized egg). In animals such as salamanders, cells will also *re*-build complex structures such as limbs, no matter where along the limb axis they are amputated, and *stop when it is complete*. While this regenerative capacity is not limitless, the basic observation is that the cells cooperate toward a specific, invariant endstate (the target morphology), from diverse starting conditions, and cease their activity when the correct pattern has been achieved. Thus, the cells do not merely perform a rote set of steps toward an emergent outcome, but modify their activity in a context-dependent manner to achieve a specific anatomical target morphology. In this, morphogenetic

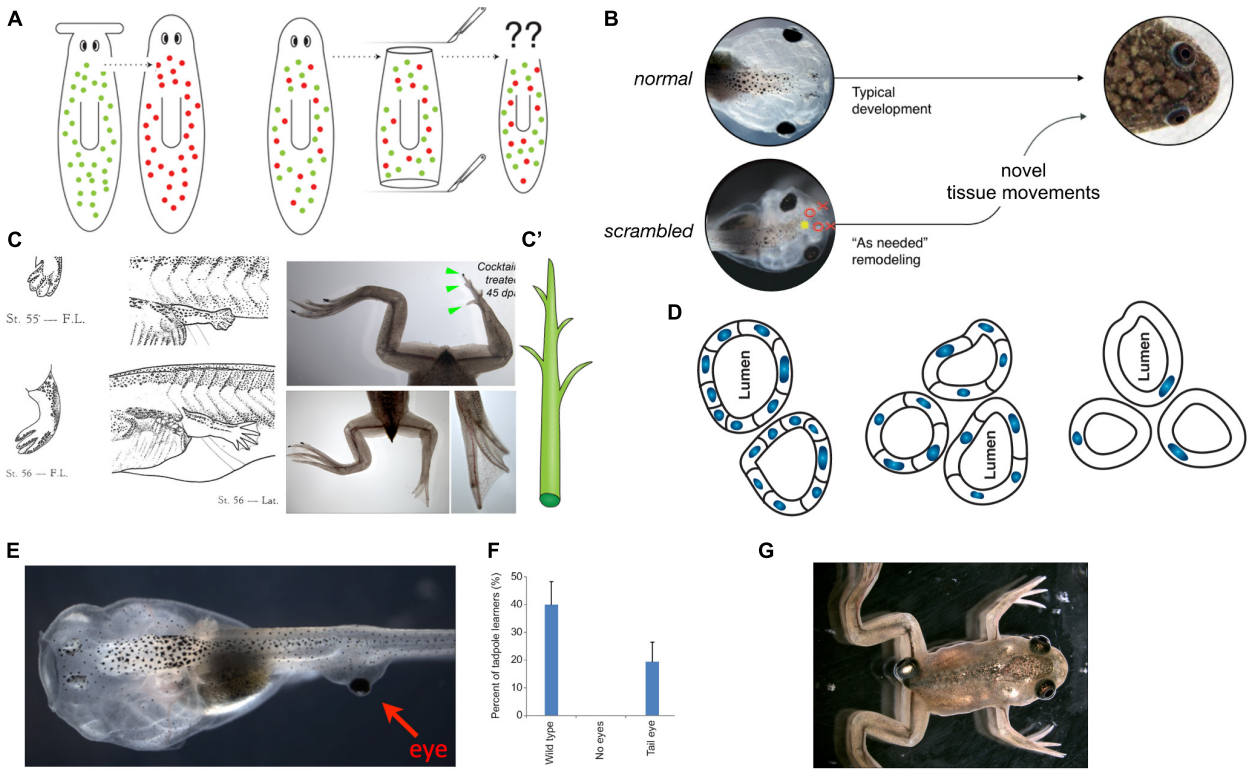


FIGURE 6 | Morphogenesis as an example of collective intelligence and plasticity. The results of complex morphogenesis are the behavior in morphospace of a collective intelligence of cells. It is essential to understand this collective intelligence because by themselves, progress in molecular genetics is insufficient. For example, despite genomic information and much pathway data on the behavior of stem cells in planarian regeneration, there are no models predicting what happens when cells from a flat-headed species are injected into a round-headed species **(A)**: what kind of head will they make, and will regeneration/remodeling ever stop, since the target morphology can never match what either set of cells expects? Development has the ability to overcome unpredictable perturbations to reach its goals in morphospace: tadpoles made with scrambled positions of craniofacial organs can make normal frogs **(B)** because the tissues will move from their abnormal starting positions in novel ways until a correct frog face is achieved (Vandenberg et al., 2012). This illustrates that the genetics seeds the development of hardware executing not an invariant set of movements but rather an error minimization (homeostatic) loop with reference to a stored anatomical setpoint (target morphology). The paths through morphospace are not unique, illustrated by the fact that when frog legs are induced to regenerate **(C)**, the intermediate stages are not like the developmental path of limb development (forming a paddle and using programmed cell death to separate the digits) but rather like a plant **(C')**, in which a central core gives rise to digits growing as offshoots (green arrowheads) which nevertheless ends up being a very normal-looking frog leg (Tseng and Levin, 2013). **(D)** The plasticity extends across levels: when newt cells are made very large by induced polyploidy, they not only adjust the number of cells that work together to build kidney tubules with correct lumen diameter, but can call up a completely different molecular mechanism (cytoskeletal bending instead of cell:cell communication) to make a tubule consisting in cross-section of just 1 cell wrapped around itself; this illustrates intelligence of the collective, as it creatively deploys diverse lower-level modules to solve novel problems. The plasticity is not only structural but functional: when tadpoles are created to **(E)** have eyes on their tails (instead of in their heads), the animals can see very well (Blackiston and Levin, 2013), as revealed by their performance in visual learning paradigms **(F)**. Such eyes are also competent modules: they first form correctly despite their aberrant neighbors (muscle, instead of brain), then put out optic nerves which they connect to the nearby spinal cord, and later they ignore the programmed cell death of the tail, riding it backward to end up on the posterior of the frog **(G)**. All of this reveals the remarkable multi-scale competency of the system which can adapt to novel configurations on the fly, not requiring evolutionary timescales for adaptive functionality (and providing important buffering for mutations that make changes whose disruptive consequences are hidden from selection by the ability of modules to get their job done despite changes in their environment). Panels **(A,C',D)** are courtesy of Peregrine Creative. Panel **(C)** is from Xenbase and Aisun Tseng. Panels **(E–G)** are courtesy of Douglas Blackiston. Panel **(B)** is used with permission from Vandenberg et al. (2012), and courtesy of Erin Switzer.

systems meet James' test for minimal mentality: "fixed ends with varying means" (James, 1890).

For example, tadpoles turn into frogs by rearranging their craniofacial structures: the eyes, nostrils, and jaws move as needed to turn a tadpole face into a frog face (**Figure 6B**). Guided by the hypothesis that this was not a hardwired but an intelligent process that could reach its goal despite novel challenges, we made tadpoles in which these organs were in the wrong positions—so-called Picasso Tadpoles (Vandenberg et al., 2012). Amazingly, they tend to turn into largely normal frogs because the craniofacial organs move in novel, abnormal paths [sometimes overshooting and needing to return a bit (Pinet et al., 2019)] and stop *when they get to the correct frog face positions*. Similarly, frog legs that are artificially induced to regenerate create a correct final form but not *via* the normal developmental steps (Tseng and Levin, 2013). Students who encounter such phenomena and have not yet been inoculated with the belief that molecular biology is a privileged level of explanation (Noble, 2012) ask the obvious (and proper) question: how does it know what a correct face or leg shape is?

Examples of remodeling, regulative development (e.g., embryos that can be cut in half and produce normal monozygotic twins), and regeneration, ideally illustrate the goal-directed nature of cellular collectives. They pursue specific anatomical states that are much larger than any individual cells and solve problems in morphospace in a context-sensitive manner—any swarm of miniature robots that could do this would be called a triumph of collective intelligence in the engineering field. Guided by the TAME framework, two questions come within reach. First, how does the collective measure current state and store the information about the correct target morphology? Second, if morphogenesis is not at the clockwork level on the continuum of persuadability but perhaps at that of the thermostat, could it be possible to re-write the setpoint without rewiring the machine (i.e., in the context of a wild-type genome)?

Pattern Memory: A Key Component of Homeostatic Loops

Deer farmers have long known of trophic memory: wounds made on a branched antler structure in 1 year, will result in ectopic tines growing *at that same location* in subsequent years, long after the original rack of antlers has fallen off (Bubenik and Pavlansky, 1965; Bubenik, 1966; Lobo et al., 2014). This process requires cells at the growth plate in the scalp to sense, and remember for months, the location of a transient damage event within a stereotypical branched structure, and reproduce it in subsequent years by over-riding the wild-type stereotypical growth patterns of cells, instead guiding them to a novel outcome. This is an example of experience-dependent, re-writable pattern memory, in which the target morphology (the setpoint for anatomical homeostasis) is re-written within standard hardware.

Planarian flatworms can be cut into multiple pieces, and each fragment regenerates precisely what is missing at each location (and re-scales the remaining tissue as needed) to make a perfect little worm (Cebrià et al., 2018). Some species of planaria have an incredibly messy genome—they are mixoploid due to their method of reproduction: fission and regeneration, which propagates any mutations that don't kill the stem cell and expands it throughout the lineage [reviewed in Fields et al. (2020)]. Despite the divergence of genomic information, the worms are champion regenerators, with near 100% fidelity of anatomical structure. Recent data have identified one set of mechanisms mediating the ability of the cells to make, for example, the correct number of heads: a standing bioelectrical distribution across the tissue, generated by ion channels and propagated by electrical synapses known as gap junctions (**Figures 7A–D**). Manipulation of the normal voltage pattern by targeting the gap junctions (Sordillo and Bargmann, 2021) or ion channels can give rise to planaria with one, two, or 0 heads, or heads with shape (and brain shape) resembling other extant species of planaria (Emmons-Bell et al., 2015; Sullivan et al., 2016). Remarkably, the worms with abnormal head number are *permanently* altered to this pattern, despite their wild-type genetics: cut into pieces with no further manipulations, the pieces continue to regenerate with abnormal head number (Oviedo et al., 2010; Durant et al., 2017). Thus, much like the optogenetic techniques used to incept false behavioral memories into brains (Vetere et al., 2019), modulation of transient bioelectric state is a conserved mechanism by which false pattern memories can be re-written into the genetically-specified electrical circuits of a living animal.

Multi-Scale Competency of Growth and Form

A key feature of morphogenesis is that diverse underlying molecular mechanisms can be deployed to reach the same large-scale goal. This plasticity and coarse-graining over subunits' states is a hallmark of collective cognition, and is also well known in neuroscience (Prinz et al., 2004; Otopalik et al., 2017). Newt kidney tubules normally have a lumen of a specific size and are made up (in cross section) of 8–10 cells (Fankhauser, 1945a,b). When the cell size is experimentally enlarged, the same tubules are made of a smaller number of the bigger cells. Even more remarkable than the scaling of the cell number to unexpected size changes (on an ontogenetic, not evolutionary, timescale) is the fact that if the cells are made really huge, *just one cell* wraps around itself and still makes a proper lumen (**Figure 6D**). Instead of the typical cell-cell interactions that coordinate tubule formation, cytoskeletal deformations within one cell can be deployed to achieve the same end result. As in the brain, the levels of organization exhibit significant autonomy in the details of their molecular activity but are harnessed toward an invariant system-level outcome.

Specific Parallels Between Morphogenesis and Basal Cognition

The plasticity of morphogenesis is significantly isomorphic to that of brains and behavior because the communication dynamics that scale individual neural cells into a coherent Self are ones that evolution honed long before brains appeared, in the context of morphogenic control (Fields et al., 2020), and before that, in metabolic control in bacterial biofilms (Prindle et al., 2015; Liu et al., 2017; Martinez-Corral et al., 2019; Yang et al., 2020). Each genome specifies cellular hardware that implements signaling circuits with a robust, reliable default "inborn" morphology—just as genomes give rise to brain circuits that drive instinctual behavior in species that can build nests and do other complex things with no training. However, evolution selected for hardware that can be reprogrammed by experiences, in addition to its robust default functional modes—in body structure, as well as in brain-driven behavior. Many of the brain's special features are to be found, unsurprisingly, in other forms outside the central nervous system. For example, mirror neurons and somatotopic representation are seen in limbs' response to injury, where the type and site of damage to one limb can be read out within 30 s from imaging the opposite, un-injured limbs (Busse et al., 2018). **Table 2** shows the many parallels between morphogenetic and cognitive systems.

Not Just Philosophy: Why These Parallels Matter

The view of anatomical homeostasis as a collective intelligence is not a neutral philosophical viewpoint—it makes strong

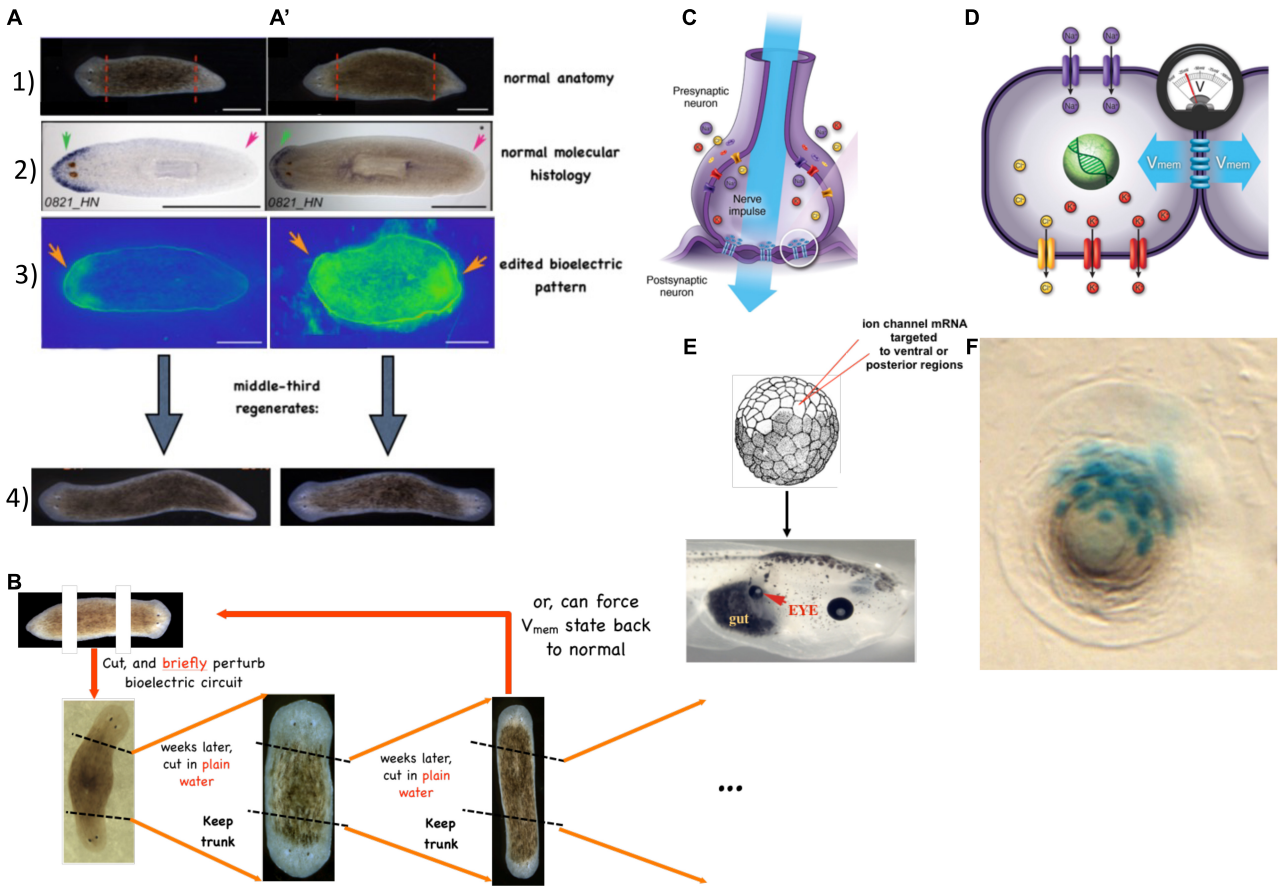


FIGURE 7 | Bioelectrical pattern memories. Planarian fragments reliably regenerate whatever is missing, and stop when a correct worm is complete. Normal planaria **(A)** have 1 head and 1 tail **(A-1)**, expression of anterior genes in the head **(A-2)**, and a standing pattern of resting potential that depolarized on the end that should make a head [**(A-3)**, revealed by voltage-reporting fluorescent dye, depolarized region marked with orange arrowhead]. When a middle portion is amputated, it regenerates to a correct 1-headed worm **(A-4)**. It is possible to edit the information structure which encodes the target morphology (shape to which the fragments will regenerate). In worms that are anatomically normal **(A′-1)**, with normal gene expression **(A′-2)**, their bioelectric pattern can be altered in place [**(A′-3)**, orange arrowheads mark the two depolarized ends] using ion channel-targeting drugs or RNAi. The result, after injury, will be a fully viable 2-headed worm **(A′-4)**. Importantly, the pattern shown in panel **(A′-3)** is not a voltage map of the final 2-headed worm: it's a map of a 1-headed animal before cutting, which already has the induced false memory indicating that a correct worm should have 2 heads. In other words, the bioelectrics can diverge from the current state—it is not simply a readout of what the anatomy is doing now, but an orthogonal information medium that is used to guide future changes of anatomy. This information is latent, only guiding the cellular collective's anatomical homeostasis activity after injury. Thus it is also a basal example of counterfactual representation, referring to what should happen *if* an injury occurs, not what is happening now. Such changes to the bioelectric target morphology are true memories because they are re-writable but also long-term stable: if cut again, in water with no more channel-perturbing reagents, multiple rounds of regeneration of a genetically wild-type worm continue to give rise to 2-headed forms **(B)**, which can be re-set back to normal by a different bioelectric perturbation (Oviedo et al., 2010). The control of morphology by bioelectric patterns is mediated as in the brain **(C)** by cells which have ion channels that set resting potential across the membrane (V_{mem}) and propagate those states in computational networks to their neighbors, *via* electrical synapses known as gap junctions. All cells, not just neurons **(D)** do this, and bioelectric signaling is an ancient information processing modality that pre-dates neurons and brains (Fields et al., 2020; Levin, 2021a). The ability of voltage states to functionally specify modular anatomy is seen when an ion channel is used to set membrane voltage in endodermal cells fated to be gut, to an eye-like bioelectric prepattern **(E)**, which then create an eye on the gut (red arrowhead) (Pai et al., 2012). This phenomenon has 2 levels of instruction **(F)**: in addition to our use of voltage to instruct shape at the organ level (not micromanaging individual eye components), the ion channel mRNA-injected cells (cyan β-galactose marker) further instruct their neighbors (brown cells) to participate in forming this ectopic lens. Images in panels **(C,D)** are courtesy of Peregrine Creative. Images in panels **(A,A′)** are taken with permission from Durant et al. (2017). Embryo image in panel **(E)** is from Xenbase. Panel **(F)** is used with permission from Zahn et al. (2017).

predictions, some of which have already borne fruit. It led to the discovery of reprogrammable head number in planaria (Nogi and Levin, 2005) and of pre-neural roles for serotonin (Fukumoto et al., 2005a,b). It explains the teratogenicity for pre-neural exposure to ion channel or neurotransmitter drugs (Hernandez-Diaz and Levin, 2014), the patterning defects observed in human channelopathies in addition to the neurological phenotypes [reviewed in Srivastava et al. (2020)], and the utility of gap junction blockers as general anesthetics.

Prediction derived from the conservation and scaling hypotheses of TAME can be tested *via* bioinformatics. Significant and specific overlap are predicted for genes involved in morphogenesis and cognition (categories of memory and learning). This is already known for ion

TABLE 2 | isomorphism between cognition and pattern formation.

Cognitive concept	Morphogenetic concept
Patterns of activation across neural networks processing information	Differential patterns of V_{mem} across tissue formed by propagation of bioelectric states through gap junction synapses.
Local field potential (EEG)	V_{mem} distribution of cell group
Intrinsic plasticity	Change of ion channel expression based on V_{mem} levels
Synaptic plasticity	Change of cell:cell connectivity *via* V_{mem}'s regulation of gap junctional connectivity
Activity-dependent transcriptional changes	Bioelectric signals' regulating gene expression during patterning
Neuromodulation, and neurotransmitters controlled by electrical dynamics to regulate genes in neurons	Developmental (pre-nervous) signaling *via* the same neurotransmitters (e.g., serotonin) moving under control of bioelectrical gradients to regulate second messenger pathways and gene expression.
Direct transmission	Cell:cell sharing of voltage *via* nanotubes or gap junctions
Volume transmission	Cell:cell communication *via* ion levels outside the membrane or voltage-dependent neurotransmitter release
Synaptic Vesicles	Exosomes
Sensitization	Cells become sensitized stimuli, such as for example to BMP antagonists during development
Functional lateralization	Left-right asymmetry of body organs
Taste and olfactory perception	Morphogenetic signaling by diffusible biochemical ligands
Activity-dependent modification of CNS	Control of anatomy by bioelectric signaling within those same cells
Critical plasticity periods	Competency windows for developmental induction events
Inborn behaviors (instincts)	Emergent morphogenetic cascades as "default" outcomes of a genetically-specified bioelectric hardware—hardwired patterning programs (mosaic development
Voluntary movement	Remodeling, regeneration, metamorphosis
Memory	Short range: epigenetic cell memory Medium range: Regeneration of specific body organs. Long range: Morphological homeostasis over decades as individual cells senesce; altering basic body anatomy in planaria by direct manipulation of bioelectric circuit
Counterfactual memories	Ability of 1-headed planarian bodies to store bioelectric patterns indicative of 1-headed or 2-headed forms, which are latent memories that become instructive upon damage to the organism.
Perceptual Bistability	Cryptic Planaria, induced by gap-junctional disruption, fragments of which stochastically regenerate as 1-headed or 2-headed forms, shifting between two different bioelectrical representations of a target morphology (pattern memory).
Edge detection in retina	Sharp boundaries between regions of different V_{mem} induce downstream gene expression and morphogenetic outcomes
Pattern completion ability of neural networks (e.g., attractor nets)	Regeneration of missing parts in partial fragments (e.g., planaria, salamander appendages, etc.)
Forgetting	Degradation of target morphology setpoint information leading to cancer and loss of regenerative ability
Addiction	Dependency on cellular signals, such as nerve addiction in limb regeneration and cancer addiction to specific molecules.
Encoding	Representation of patterning goal states by bioelectric properties of tissue
Visual system feature detection	Organ-level monitoring of body configuration and detection of specific boundaries by tissue (such as the V_{mem} boundary that drives brain morphogenesis)
Holographic (distributed) storage	Any small piece of a planarian remembers the correct pattern (even if it has been re-written)
Behavioral plasticity	Regulative developmental programs and regenerative capacity
Self-modeling	Representations of current and future morphogenetic states by bioelectric patterns such as the planarian prepattern or the bioelectric face pattern in vertebrates
Goal-seeking	Embryogenesis and regeneration work toward a specific target configuration despite perturbations
Adaptivity and Intelligence	Morphological rearrangements carry out novel, not hardwired, movements to reach the same anatomical configuration despite unpredictable initial starting state
Age-dependent cognitive decline	Age-dependent loss of regenerative ability
Top-down control	Place conditioning for drug effects—top-down control of signaling pathways

Possible mapping of concepts in cognitive neuroscience to examples in pattern formation.

channels, connexin (gap junction) genes, and neurotransmitter machinery, but TAME predicts a widespread re-use of the same molecular machinery. Cell-cell communication and cellular stress pathways should be involved in internal conflict between psychological modules (Reinders et al., 2019) and social behavior, while memory genes should be

identified in genetic investigations of cancer, regeneration, and embryogenesis.

Another key prediction that remains to be tested (ongoing in our lab) is trainability of morphogenesis. The collective intelligence of tissues could be sophisticated enough to be trainable *via* reinforcement learning for specific morphological outcomes. Learning has been suggested by clinical data in the heart (Zoghi, 2004), bone (Turner et al., 2002; Spencer and Genever, 2003), and pancreas (Goel and Mehta, 2013). It is predicted that using rewards and punishments (with nutrients/endorphins and shock), not micromanagement of pathway hardware, could be a path to anatomical control in clinical settings, whether for morphology or for gene expression (Biswas et al., 2021). This would have massive implications for regenerative medicine, because the complexity barrier prevents advances such as genomic editing from impacting e.g., limb regeneration in the foreseeable future. The same reasons for which we would train a rat for a specific behavior, rather than control all of the relevant neurons to force it to do it like a puppet, explain why the direct control of molecular hardware is a far more difficult biomedical path than understanding the sets of stimuli that could motivate tissues to build a specific desired structure.

The key lesson of computer science has been that even with hardware we understand (if we built it ourselves), it is much more efficient and powerful to understand the software and evince desired outcomes by the appropriate stimulation and signaling, not physical rewiring. If the hardware is reprogrammable (and it is here argued that much of the biological hardware meets this transition), one can offload much of the complexity onto the system itself, taking advantage of whatever competence the sub-modules have. Indeed, neuroscience itself may benefit from cracking a simpler version of the problem, in the sense of neural decoding, done first in non-neural tissues.

Non-neural Bioelectricity: What Bodies Think About

The hardware of the brain consists of ion channels which set the cells' electrical state, and controllable synapses (e.g., gap junctions) which can propagate those states across the network. This machinery, including the neurotransmitters that eventually transduce these computations into transcriptional and other cell behaviors, is in fact highly conserved and present in all cells, from the time of fertilization (**Figures 7C,D**). A major difference between neural and non-neural bioelectricity is the time constant with which it acts [brains speed up the system into millisecond scales, while developmental voltage changes occur in minutes or hours (Harris, 2021; Levin, 2021a)]. Key aspects of this system in any tissue that enable it to support flexible software include the fact that both ion channels and gap junctions are themselves voltage sensitive—in effect, they are transistors (voltage-gated current conductances). This enables evolution to exploit the laws of physics to rapidly generate very complex circuits with positive (memory) and negative (robustness) feedback (Law and Levin, 2015; Cervera et al., 2018, 2019a,b, 2020a). The fact that a transient voltage state passing through a cell can set off a cycle of progressive depolarization (like an action potential) or gap junctional (GJ) closure means that such circuits readily form dynamical systems memories which can store different information and change their computational behavior without changing the hardware (i.e., not requiring new channels or gap junctions) (Pietak and Levin, 2017); this is obvious in the action potential propagations in neural networks but is rarely thought about in development. It should be noted that there are many additional biophysical modalities, such as parahormones, volume conduction, biomechanics (strain and other forces), cytoskeletal dynamics, and perhaps even quantum coherence events that could likewise play interesting roles. These are not discussed here only due to length limitations; instead, we are focusing on the bioelectric mechanisms as one particularly illustrative example of how evolution exploits physics for computation and cognition.

Consistent with its proposed role, slowly-changing resting potentials serve as instructive patterns guiding embryogenesis, regeneration, and cancer suppression (Bates, 2015; Levin et al., 2017; McLaughlin and Levin, 2018). In addition to the pattern memories encoded electrically in planaria (discussed above), bioelectric prepatterns have also been shown to dictate the morphogenesis of the face, limbs, and brain, and function in determining primary body axes, size, and organ identity [reviewed in Levin and Martyniuk (2018)]. One of the most interesting aspects of developmental bioelectricity is its modular nature: very simple voltage states trigger complex, downstream patterning cascades. As in the brain, modularity goes hand-in-hand with pattern completion: the ability of such networks to provide entire behaviors from partial inputs. For example, **Figure 7F** shows how a few cells transduced with an ion channel that sets them into a "make the eye here" trigger recruit their neighbors, in any region of the body, to fulfill the purpose of the subroutine call and create an eye. Such modularity makes it very easy for evolution to develop novel patterns by re-using powerful triggers. Moreover, as do brains, tissues use bioelectric circuits to implement pattern memories that set the target morphology for anatomical homeostasis (as seen in the planarian examples above). This reveals the non-neural material substrate that stores the information in cellular collectives, which is a distributed, dynamic, re-writable form of storage that parallels recent discoveries of how group knowledge is stored in larger-scale agents such as animal swarms (Thierry et al., 1995; Couzin et al., 2002). Finally, bioelectric domains (Pai et al., 2017, 2018; Pitcairn et al., 2017; McNamara et al., 2019, 2020) set the borders for groups of cells that are going to complete a specific morphogenetic outcome—a system-level process like "make an eye." They define the spatio-temporal borders of the modular activity, and suggest a powerful model for how Selves scale in general.

A Bioelectric Model of the Scaling of the Self

Gap junctional connections between cells provide an interesting case study for how the borders of the Self can expand or contract, in the case of a morphogenetic collective intelligence (**Figure 8**). Crucially, gap junctions [and gap junctions extended by tunneling nanotubes (Wang et al., 2010; Ariazi et al., 2017)]

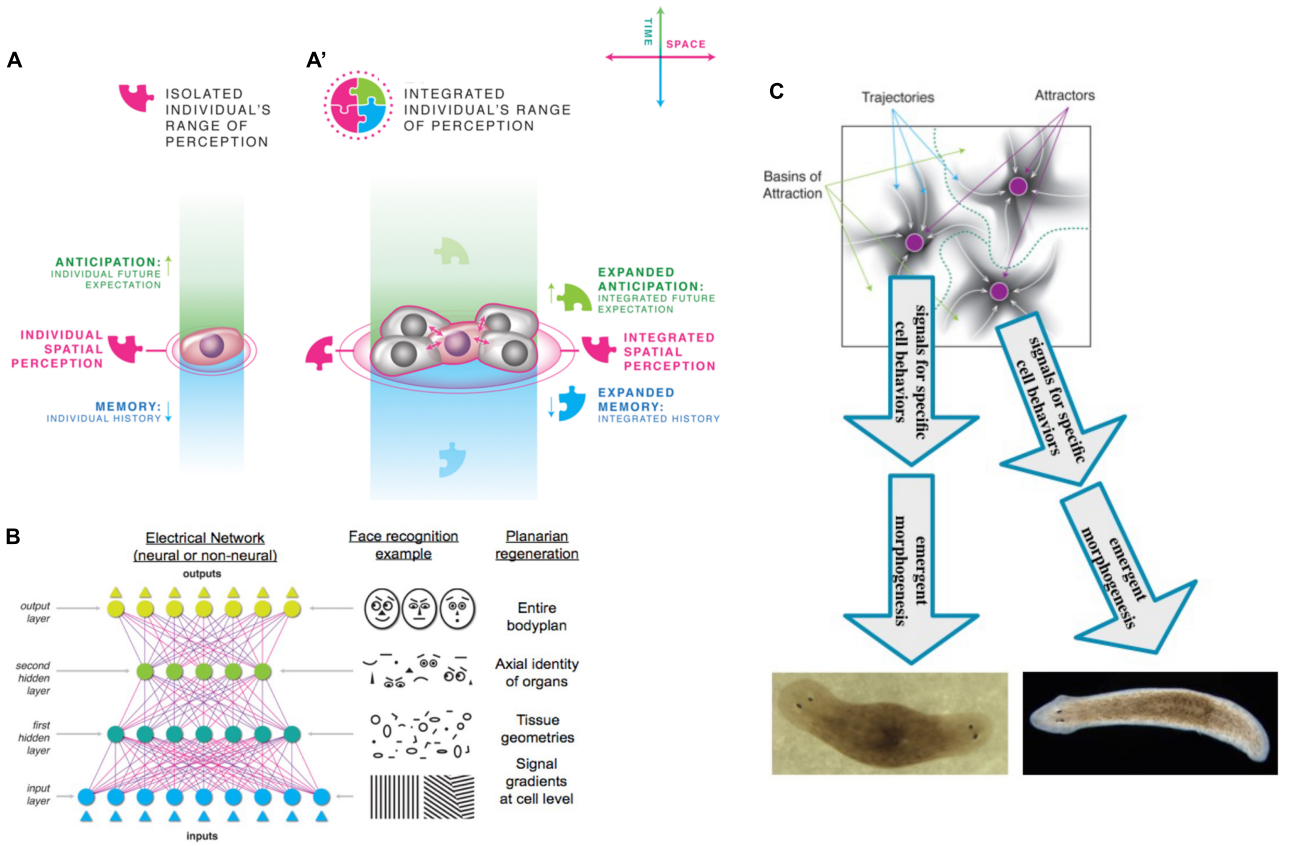


FIGURE 8 | Scaling of computation in cells. Individual cells **(A)** have a degree of computational capacity consisting of the ability to sense local microenvironment, and some memory and ability to anticipate into the future. When assembling into networks **(A')**, tissues acquire the ability to sense and act at greater spatial distance, as well as gain larger capacity for memory and prediction *via* greater computational capacity. As neural networks use hidden layers to abstract patterns in data and recognize meso-scale features **(B)**, tissue networks gain the capacity to represent information larger than the molecular and cell level: each cell's activity (differentiation, migration, etc.) can be the result of other layers of cells processing information about current and past states, enabling decision-making with respect to tissue, organ, or whole organism-scale anatomy. **(C)** Much as some neural networks store individual memories as attractors in their state space, bioelectric circuits' attractors function as pattern memories, triggering cells to execute behaviors that implement anatomical outcomes like number and location of heads in planaria. Images courtesy of Jeremy Guay of Peregrine Creative.

enable a kind of cellular parabiosis—a regulated fusion between cells that enables lateral inheritance of physiological information, which speeds up processing in the same way that lateral gene inheritance potentiates change on evolutionary timescales. The following is a case study hypothesizing one way in which evolution solves the many-into-one problem (how competent smaller Selves bind into an emergent higher Self), and how this process can break down leading to a reversal (shrinking) of the Self boundary (summarized in **Table 3**).

Single cells (e.g., the protozoan *Lacrymaria olor*) are very competent in handling morphological, physiological, and behavioral goals on the scale of one cell. When connected to each other *via* gap junctions, as in metazoan embryos, several things happen (much of which is familiar to neuroscientists and workers in machine learning in terms of the benefits of neural networks) that lead to the creation of a Self with a new, larger cognitive boundary. First, when cells join into an electrochemical network, they can now sense events, and act, on a much larger physical "radius of concern" than a single cell. Moreover, the network can now integrate information coming from spatially disparate regions in complex ways that result in activity in other spatial regions. Second, the network has much more computational power than any of its individual cells (nodes), providing an IQ boost for the newly formed Self. In such networks, Hebbian dynamics on the electrical synapse (GJ) can provide association between action in one location and reward in another, which enables the system to support credit assignment at the level of the larger individual.

The third consequence of GJ connectivity is the partial dissolution of informational boundaries between the subunits. GJ-mediated signals are unique because they give each cell immediate access to the internal milieu of other cells. A conventional secreted biochemical signal arrives from the outside, and when it triggers cell receptors on the surface, the cell clearly knows that this information originated externally (and can be attended to, ignored, etc.)—it is easy to maintain boundary between Self and world. However, imagine a signal like a calcium spike originating in a cell due to some damage

TABLE 3 | An example of the scaling of cognition.

- Each Self has a cognitive capacity defined by the spatial, temporal, and complexity metrics on the goals it can possibly pursue.
- Biological Selves scale up by cells' joining into computational networks that can pursue larger-scale (anatomical, not just metabolic) goals.
- Networks increase the spatial reach of sensing and actuation, and increase the computational capacity which allows scaling up of goals and of the states that can induce stress.
- Bodies consist of components which are themselves competent (goal-seeking modules that navigate their own spaces) and can achieve specific outcomes despite perturbations and changing conditions.
- Gap junctions are a unique scaling mechanism which, by linking cells' internal milieus, wipes ownership information on signaling molecules. This partially erases the informational identity of the cellular subunits, driving up cooperation and resulting in novel tissue and organ-level Selves with morphological-scale goals.
- Bioelectric networks underlie the computations of cell collectives at the tissue, organ, and organism scale, propagating stress information, state sensing, and morphogenetic instructive cues over larger areas.
- Selves can dissociate (scale down), as occurs in cancer, by shrinking the computational boundaries of some subunits that de-couple from the network.

stimulus for example. When that calcium propagates onto the GJ-coupled neighbor, there are no metadata on that signal marking its origin; the recipient cell only knows that a calcium transient occurred, and cannot tell that this information does not belong to it. The downstream effects of this second messenger event are a kind of false memory for the recipient cell, but a true memory for the collective network of the stimulus that occurred in one part of the individual. This wiping of ownership information for GJ signals as they propagate through the network is critical to enabling a partial "mind meld" between the cells: keeping identity (in terms of distinct individual history of physiological states—memory) becomes very difficult, as small informational molecules propagate and mix within the network. Thus, this property of GJ coupling promotes the creation of a larger Self by partially erasing the mnemic boundaries between the parts which might impair their ability to work toward a common goal. This is a key part of the scaling of the Self by enlarging toward common goals—not by micromanagement, but by bringing multiple subunits into the same goal-directed loop by tightly coupling the sensing, memory, and action steps in a syncytium where all activity is bound toward a system-level teleonomic process. When individual identities are blurred in favor of longer time-scale, larger computations in tissues, small-horizon (myopic) action in individual cells (e.g., cancer cells' temporary gains followed by maladaptive death of the host) leads to a more adaptive longer-term future as a healthy organism. In effect, this builds up long-term collective rationality from the action of short-sighted irrational agents (Sasaki and Biro, 2017; Berdahl et al., 2018).

It is important to note that part of this story has already been empirically tested in assays that reveal the shrinking as well as the expansion of the Self boundary (**Figure 9**). One implication of these hypotheses is that the binding process can break down. Indeed this occurs in cancer, where oncogene expression and carcinogen exposure leads to a closure of GJs (Vine and Bertram, 2002; Leithe et al., 2006). The consequence of this is transformation to cancer, where cells revert to their ancient unicellular selves (Levin, 2021b)—shrinking their computational boundaries and treating the rest of the body as external environment. The cells migrate at will and proliferate as much as they can, fulfilling their cell-level goals—metastasis [but also sometimes attempting to, poorly, reboot their multicellularity and make tumors (Egeblad et al., 2010)]. The model implies that this phenotype can be reverted by artificially managing the bioelectric connections between a cell and its neighbors. Indeed, recent data show that managing this connectivity can override default genetically-determined states, inducing metastatic melanoma in a perfectly wild-type background (Blackiston et al., 2011) or suppressing tumorigenesis induced by strong oncogenes like p53 or KRAS mutations (Chernet and Levin, 2013a,b). The focus on physiological connectivity (information dynamics)—the software—is consistent with the observed facts that genetic alterations (hardware) are not necessary to either induce or revert cancer [reviewed in Chernet and Levin (2013a)].

All these dynamics lead to a few interesting consequences. GJ-mediated communications are not merely conversations (in the way that external signaling is)—they are binding, in the sense that once a GJ is open, a cell is subject to whatever comes in from the neighbor. In the same sense, having a synapse makes one vulnerable to the state of neighbors. GJs spread (dilute) the pain of depolarization, but at the same time give a cell's neighbors the power to change its state. Compatible with the proposal that the magnitude of a Self is the scale and complexity of states by which it can be stressed, connections by tunable, dynamic GJs greatly expand the spatial, temporal, and complexity of things that can irritate cells; complex events from far away can now percolate into a cell *via* non-linear GJ paths through the network, and enabling the drive to minimize such events now necessarily involves homeostatic activity of goal states, sensing, and activity on a much larger scale. Stress signals, propagating through such networks, incentivize other regions of the tissue to act cooperatively in response to *distant* events by harnessing their selfish drive to reduce their own stress. This facilitates the coherent, system-level response to perturbations beyond their local consequences, and gives rise to larger Selves able to react coherently to stressful departures from more complex, spatially-distributed allostatic setpoints. For example, whereas a solitary cell might be stressed (and react to) abnormal local pH, cells that are part of a transplanted salamander limb will be induced to a more grandiose activity: they will change the number of fingers they produce to be correct relative to the limb's new position in the host's body (Ruud, 1929), a decision that involves large-scale sensing, decision-making, and control.

A fourth consequence of the coupling is that cooperation in general is greatly enhanced. In the game theory sense, it is impossible to cheat against your neighbor if you are physiologically coupled. Any positive or negative effects of a cell's actions toward the neighboring cell are immediately propagated back to it, in effect producing one individual in which the parts cannot "defect" against each other. This dynamic suggests an interesting twist on Prisoners' Dilemma models

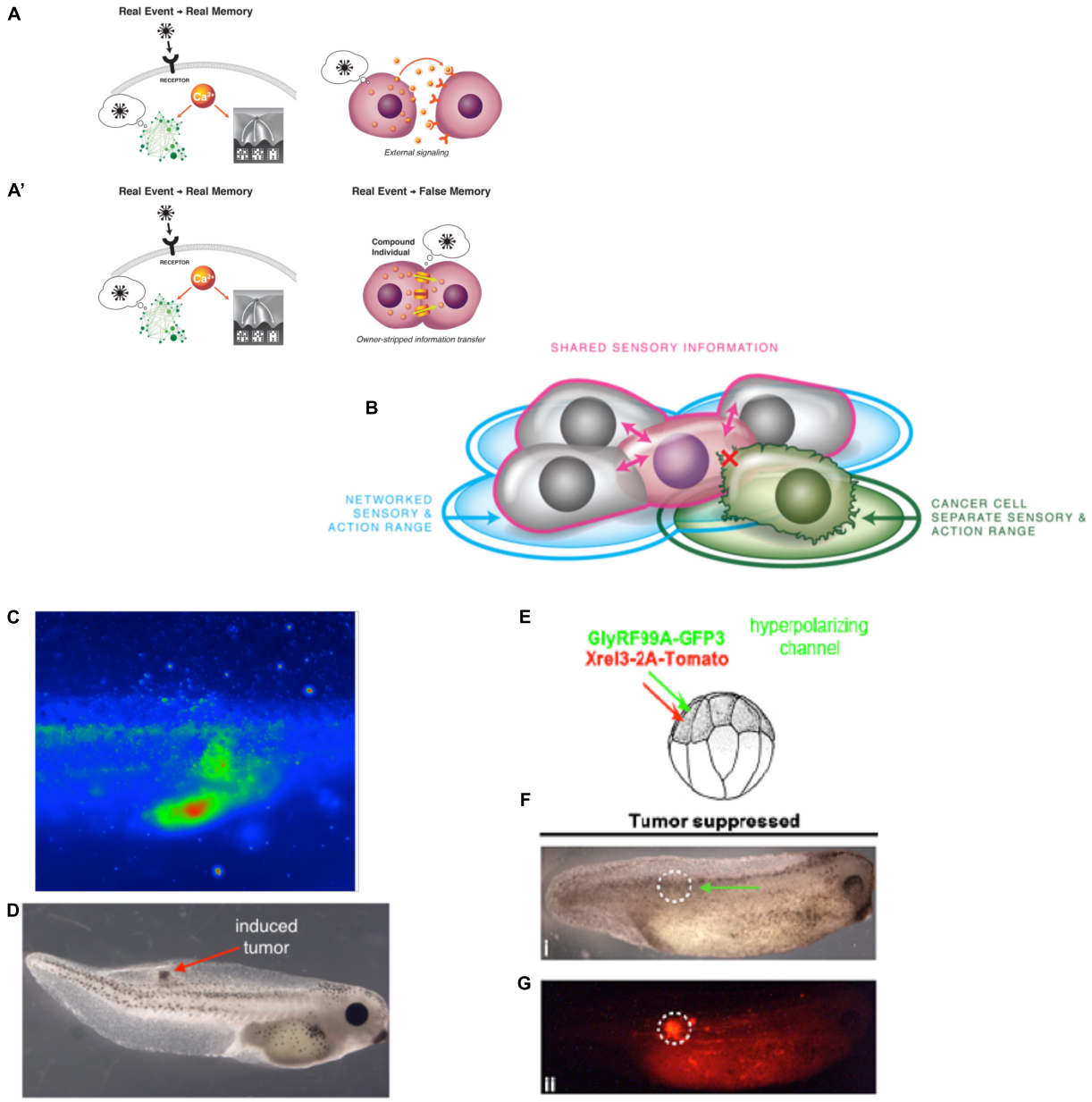


FIGURE 9 | Gap junctions and the cellular collective. Communication *via* diffusible and biomechanical signals can be sensed by receptors at the membrane as messages coming from the outside of a cell **(A)**. In contrast, cells coupled by gap junctions enable signals to pass directly from one cell's internal milieu into another. This forms a partial syncytium which helps erase informational boundaries between cells, as memory molecules (results of pathway dynamics) propagate across such cell groups without metadata on which cell originated them. The versatile gating of GJ synapses allows the formation of multicellular Selves that own memories of physiological past events at the tissue level (not just individual cells') and support larger target patterns, enabling them to cooperate to make complex organs **(B)**. This process can break down: when oncogenes are expressed in tadpoles, voltage dye imaging **(C)** reveals the abnormal voltage state of cells that are disconnected bioelectrically from their neighbors, reverting to an ancient unicellular state (metastasis) that treats the rest of the body as external environment and grows out of control as tumors **(D)**. This process can be prevented (Chernet and Levin, 2013a,b; Chernet et al., 2016) by artificially regulating their bioelectric state [e.g., co-injecting a hyperpolarizing channel with the oncogene, **(E)**]. In this case the tissue forms normally [**(F)**, green arrow], despite the very strong presence of the oncogene [**(G)**, red label]. This illustrates the instructive capacity of bioelectric networks to dominate single cell and genetic states to control large-scale tissue outcomes. Panels **(A,A′,B)** courtesy of Jeremy Guay of Peregrine Creative. Panels **(C–D)** are used with permission from Chernet and Levin (2013a). Panels **(E–G)** used with permission from Chernet and Levin (2013b).

in which the number of agents is not fixed, because they have the options of Cooperate, Defect, Merge, and Split (we are currently analyzing such models). Specifically, merging with another agent creates an important dimensionality reduction (because defection is no longer an option); this not only changes the calculus of game theory as applied to biological

interactions, but also the action space itself. These dynamics take place on a developmental timescale, complementing the rich existing literature on game theory in evolution (Maynard Smith and Szathmáry, 1995; Maynard Smith, 1999; McEvoy, 2009; Pacheco et al., 2014).

Indeed, the smaller and larger agents' traversal of their various spaces provides a way to think about how smaller agents' (cell-level) simple homeostatic loops can scale up into large, organ-level anatomical homeostatic loops. Prentner recently showed how agents build up spatial models of their worlds by taking actions that nullify changes in their experience (Prentner, 2019). Working to nullify changes to one's state that would otherwise be induced by the vagaries of external environment (and other agents) is the core of homeostasis—the action loops that seek to preserve important states against intervention and entropy. This is not only for physical movement (which results in a creature perceiving itself to be situated in spacetime) but also for other states in which actuation takes place *via* turning on/off specific genes, remodeling an anatomy, or opening/closing ion channels to change physiological state. An agent can notice patterns in what actions it had to take to keep in homeostasis despite various perturbations that occur, and based on that refine an internal model of some space within which it is acting. This is closely related to the surprise minimization framework (Friston, 2013; Friston et al., 2013; Friston K. et al., 2014), and suggests a straightforward sense in which larger Selves scale up to models of their world and themselves from evolutionary primitives such as metabolic homeostasis. Bioelectricity provides examples where cell-level physiological homeostats form networks that implement much larger-scale pattern memories as attractors, akin to Hopfield networks (**Figure 10**; Hopfield, 1982; Inoue, 2008; Pietak and Levin, 2017; Cervera et al., 2018, 2019a,b, 2020a). This enables all tissues to participate in the kind of pattern completion seen in neural networks—a critical capability for regenerative and developmental repair (anatomical homeostasis).

With these pieces in place, it is now possible to mechanistically visualize one specific aspect of the progressive scaling that expands the cognitive light cone. Cells with a chemical receptor can engage in predictive coding to manage their sensory experience (Friston K. et al., 2014; Friston K. J. et al., 2014; Thornton, 2017). Similarly, individual cells homeostatically maintain V_{mem} (cell membrane resting potential voltage) levels. However, cells can electrically couple *via* gap junctions to create bioelectric networks that work as a kind of virtual governor—coupled oscillators possess emergent dynamics that now maintain large, spatial *patterns* of V_{mem} against perturbation with greater stability (Pietak and Levin, 2017, 2018; Cervera et al., 2018, 2019a,b, 2020a,b; Pai et al., 2018; Manicka and Levin, 2019a). These spatial patterns serve as instructive pattern memories guiding the activity of a cell collective through anatomical morphospace toward the correct target morphology (Sullivan et al., 2016; Levin, 2021a; Pezzulo et al., 2021).

Voltage is especially interesting because each V_{mem} level in a single cell is a coarse-grained parameter, subsuming many distinct combinations of sodium, potassium, chloride, etc., levels, and many distinct open/closed states of particular ion channel proteins, which all result in the same resting potential. This can be seen as the minimal version of generalization—cells learning to respond to *classes* of events by transducing not specific ion levels or channel protein activity states but the macrovariable "voltage." This in turn enables them to repurpose existing responses for novel combinations of stimuli (e.g., familiar depolarization events caused by novel ion dynamics).

Moreover, gap junctions propagate voltage states across tissue, allowing cells to respond to events that are not local in nature (larger-scale) and to respond *en masse*. More generally, this means that the input to any group of cells is produced by the output of groups of cells—sub-networks, which can be complex and highly processed over time (not instantaneous), enabling predictive coding to manage complex states (at a distance in space and time) and not only purely local, immediate sensory data. It also means that the system is extremely modular, readily connecting diverse upstream and downstream events to each other *via* the universal adapter of bioelectric states. When this is applied to the homeostatic TOTE (test-operate-exit) loop, allowing its measurement, comparison, and action modules to be independently scaled up (across space, time, and complexity metrics), this inherently expands the cognitive light cone of a homeostatic agent to enable progressively more grandiose goals.

Crucially, all of the above-mentioned aspects of the role of generic bioelectric networks underlying the scaling of Selves are not only the products of the evolutionary process, but have many functional implications for evolution itself (forming a positive feedback loop in which rising multiscale agency potentiates the evolution of increasingly more complex versions).

EVOLUTIONARY ASPECTS

Developmental bioelectricity works alongside other modalities such as gene-regulatory networks, biomechanics, and biochemical systems. The TAME framework emphasizes that what makes it special is that it's not just another micro-mechanism that developmental biologists need to track. First, developmental bioelectrics is a unique computational layer that provides a tractable entrypoint into the informational architecture and content of the collective intelligence of morphogenesis. Second, bioelectric circuits show examples of modularity, memory, spatial integration, and generalization (abstraction over ion channel microstates)—critical aspects of understanding basal origins of cognition. Developmental bioelectricity provides a bridge between the early problem-solving of body anatomy and the more recent complexity of behavioral sophistication *via* brains. This unification of two disciplines suggests a number of hypotheses about the evolutionary path that pivoted morphogenetic control mechanisms into the cognitive capacities of behavior, and thus sheds light on how Selves arise and expand.

Somatic Bioelectrics Reveals the Origin of Complex Cognitive Systems

Developmental bioelectrics is an ancient precursor to nervous systems. Analog bioelectrical dynamics generate patterns in

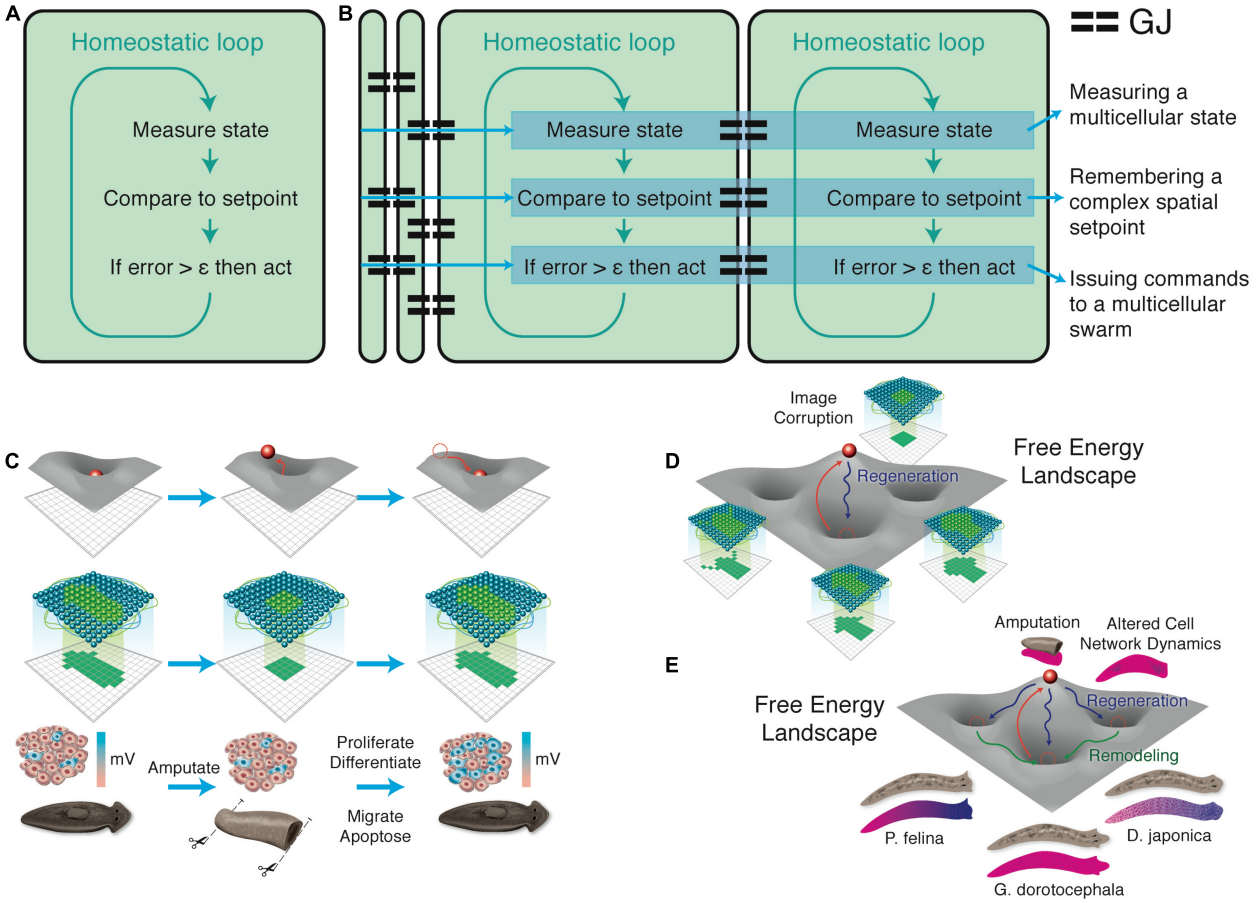


FIGURE 10 | Gap junctions scale homeostatic goals. Gap junctions are a type of connection architecture that facilitates scaling of goal states (and thus expands the cognitive cone of cellular agents). A single cell's homeostatic cycle has 3 parts: measurements, comparison to a stored setpoint, and acting *via* effectors to stay in or reach the correct region of state space **(A)**. When coupled into gap junctions **(B)**, each of these 3 components expands in size and complexity: the cell group is able to (1) measure a larger region (reacting to spatially more complex inputs, not just local conditions), (2) store a more complex setpoint pattern, and (3) act (deform, grow, etc.) at a scale that produces large-scale anatomical change. The goals of such networks readily map on to regeneration and regulative development **(C)**: dynamical systems pictures of artificial neural networks as they perform pattern completion based on partial input illustrate an energy landscape with wells corresponding to stable target morphology memories. The process of completing a correct planarian pattern from a simple fragment can be modeled in this way, perhaps with overall stress levels instantiating the free energy that the system is trying to minimize (Kuchling et al., 2020b). Such attractors correspond to different possible morphologies, and indeed the normally robust regeneration toward a single pattern **(D)** can be modified in planaria by temporarily disrupting their gap junctional network, which causes genetically un-modified worms to nevertheless build heads appropriate to other species' attractors in morphospace **(E)** (Emmons-Bell et al., 2015; Sullivan et al., 2016). Images in panels **(A,B)** courtesy of Jeremy Guay of Peregrine Creative. Images in panels **(C–E)** courtesy of Alexis Pietak.

homogenous cell sheets and coordinate information that regulates transcription and cell behaviors. Evolution first exploited this to enable cell groups to position the body configuration in developmental morphospace, long before some cells specialized to use very fast, digital spiking as neural networks for control of behavior as movement in 3-dimensional space (Fields et al., 2020). The function of nervous systems as spatial organizers operating on data from the external world (Keijzer et al., 2013) is an adaptation built upon the prior activity of bioelectric circuits in organizing the internal morphology by processing data from the internal milieu. While neural tissues electrically encode spatial information to guide movement (e.g., memory of a maze in a rat brain) by controlling muscles, bioelectric prepatterns guide the behaviors of other cell types, on slower timescales, during development, regeneration, and remodeling toward invariant, robust anatomical configurations.

Developmental bioelectricity illustrates clearly the continuous nature of properties thought to be important for cognition, and the lack of a clean line separating brainy creatures from others. On a single-cell level, even defining a "neuron" is not trivial, as most cells possess the bioelectrical machinery and a large percentage of neuronal genes are also expressed in non-neural cells (Bucher and Anderson, 2015), while neural molecular components are found in cytonemes (Huang et al., 2019). Many channel families were likely already present in the most recent unicellular ancestor (Liebeskind et al., 2015). The phylogeny of ion channels is ancient, and the appearance of context-sensitive

channels (enabling new kinds of bioelectrical feedback loops) tracks well with the appearance of complex body plans at the emergence of metazoa (Moran et al., 2015), revealing the remarkable evolutionary continuum that leads from membrane excitability in single cells to cognitive functions in advanced organisms, by way of somatic pattern control (Cook et al., 2014).

Fascinating work on bacteria has shown that prokaryotes also utilize bioelectric state for proliferation control (Stratford et al., 2019); and, paralleling the developmental data discussed above, bioelectric phenomena in bacteria scale easily from single-cell properties (Kralj et al., 2011) to the emergence of proto-bodies as bacterial biofilms. Bacterial communities use brain-like bioelectric dynamics to organize tissue-level distribution of metabolites and second messenger molecules, illustrating many of the phenomena observed in complex morphogenetic contexts, such as encoding stable information in membrane potential patterns, bistability, and spatial integration (Humphries et al., 2017; Liu et al., 2017; Larkin et al., 2018; Martinez-Corral et al., 2018; Yang et al., 2020). Not only animal lineages, but plants (Baluska and Mancuso, 2012; Volkov et al., 2019; Serre et al., 2021) use bioelectricity, as evolution frequently exploits the fact that bioelectric dynamics are a powerful and convenient medium for the computations needed to solve problems in a variety of spaces not limited to movement in 3D space.

Developmental bioelectricity helps explain how free-living cells scaled their cell-level homeostatic pathways to whole body-level anatomical homeostasis (Levin, 2019). It has long been appreciated that evolvability is potentiated by modularity—the ability to trigger complex morphogenetic cascades by a simple "master" trigger that can be re-deployed in various contexts in the body (von Dassow and Munro, 1999). Recent advances reveal that bioelectric states can form very powerful master inducers that initiate self-limiting organogenesis. For example, the action of a single ion channel can induce an eye-specific bioelectric state that creates complete eyes in gut endoderm, spinal cord, and posterior tissues (Pai et al., 2012)—locations where genetic "master regulators" like the *Pax6* transcription factor are insufficient in vertebrates (Chow et al., 1999). Likewise, misexpression of a proton pump (or a 1-h ionophore soak) to trigger bioelectric changes in an amputation wound can induce an entire 8-day cascade of building a complete tadpole tail (Adams et al., 2007; Tseng et al., 2010). This is control at the level of organ, not single cells' fate specification, and does not require the experimenter to provide all of the information needed to build the complex appendage. Thus, bioelectric states serve as effective master regulators that evolution can exploit to make modular, large-scale changes in anatomy.

Moreover, because the same V_{mem} dynamics can be produced by many different ion channel combinations, and because bioelectric states propagate their influence across tissue distance during morphogenesis (Chernet and Levin, 2014; Pai et al., 2020), evolution is free to swap out channels and explore the bioelectrical state space: simple mutations in electrogenic genes can exert very long-range, highly coordinated changes in anatomy. Indeed, the KCNH8 ion channel and a connexin were identified in the transcriptomic analysis of the evolutionary shift between two functionally different morphologies of fin structures in fish (Kang et al., 2015). The evolutionary significance of bioelectric controls can also be seen across lineages, as some viruses evolved to carry ion channel and gap junction (Vinnexin) genes that enable them to hijack bioelectric machinery used by their target cells (Shimbo et al., 1996; Hover et al., 2017).

The unique computational capabilities of bioelectric circuits likely enabled the evolution of nervous systems, as specialized adaptations of the ancient ability of all cell networks to process electrical information as pre-neural networks (Keijzer, 2015; Fields et al., 2020). A full understanding of nervous system function must involve not only its genetics and molecular biology but also the higher levels of organization comprising dynamic physiology and computations involved in memory, decision-making, and spatio-temporal integration. The same is true for the rest of the body. For example, the realization that epithelia are the generators of bioelectric information (Robinson and Messerli, 1996) suggests models in which they act like a retina wrapped around a whole embryo (and individual organs) to preprocess electrical signals into larger-scale features and compute contrast information for downstream processing (Grossberg, 1978). The investigation of somatic bioelectric states as primitive "pattern memories" and the expansion of computational science beyond neurons will enrich the understanding of cell biology at multiple scales beyond molecular mechanisms, as is currently only done with respect to the brain (Marr, 1982). Generalizing the deep concepts of multiscale neuroscience beyond neurons (Grossberg, 1978; Pezzulo and Levin, 2015; Manicka and Levin, 2019b) is necessary for a better understanding of the tissue-level decision-making that drives adaptive development and regeneration. Conversely, advances in understanding information processing in a relatively simpler anatomical context will feed back to enrich important questions in brain science, shedding light on fundamental mechanisms by which information-processing agents (cells) work collectively to accomplish unified, complex system-level outcomes. The multi-disciplinary opportunity here is not only to gain insight into the phylogeny of nervous systems and behavior, but to *erase the artificial boundaries between scientific disciplines that focus on neurons vs. the rest of the body, with the direct consequence that a more inclusive, gradualist picture emerges of the mechanisms commonly associated with cognitive Selves*.

Ion channels and gap junctions are the hardware interface to the bioelectric computational layer within living systems. Like a retina for a brain, or a keyboard for a computer, they allow transient signals to serve as inputs to memory and decision-making networks. For any given agent (cell, tissue, etc.), its bioelectrical interface is accessed by a number of potential users. First are the neighboring agents, such as other tissues, which pass on their bioelectric state during cooperative and competitive interactions in morphogenesis. There are also commensal and parasitic microbes, which have evolved to hijack such control systems to manipulate the anatomy of the host—like the naïve bacteria on planaria that can determine head number and visual system structure in flatworm regeneration (Williams et al., 2020). Moreover, the development of pharmacological, genetic, and optogenetic tools now allows human bioengineers to access bioelectrical circuits

for the control of growth and form in regenerative medicine and synthetic bioengineering contexts (Adams et al., 2013, 2014, 2016; Chernet et al., 2016; McNamara et al., 2016; Bonzanni et al., 2020). All of these manipulations can serve as catalysts, enabling an evolutionary lineage to more easily travel to regions of option space that might otherwise be separated by an energy barrier that is difficult for standard evolution to reach. In this sense, cognitive properties of developmental mechanisms help us to understand problem-solving on phylogenetic, not just ontogenetic, timescales. We next look at specific ways in which the architecture of multiscale autonomy, especially as implemented by bioelectric network mechanisms, potentiates evolution.

Multi-Scale Autonomy Potentiates the Speed of Evolution

Deterministic chaos and complexity theory have made it very clear why bottom-up control of even simple systems (e.g., 3-body problem) can be practically impossible. This inverse problem (Lobo et al., 2014)—what control signals would induce the desired change—is not only a problem for human engineers but also for adjacent biological systems such as the microbiome or a fungus that seeks to control the behavior of an ant (Hughes et al., 2016), and most of all, for other parts of a complex system (to help control itself). Evolution tackles this task by exploiting a multiscale competency architecture (MCA), where subunits making up each level of organization are themselves homeostatic agents. It's built on an extremely powerful design principle: error correction (Fields and Levin, 2017; Frank, 2019a,b).

The key aspect of biological modularity is not simply that complex subroutines can be triggered by simple signals, making it easy to recombine modules in novel ways (Schlosser and Wagner, 2004; Gerhart and Kirschner, 2007), but that these modules are also themselves sub-agents exhibiting infotaxis and socialtaxis, and solving problems in their own spaces (Vergassola et al., 2007; Gottlieb et al., 2013; Karpas et al., 2017). When an eye primordium appears in the wrong place (e.g., a tadpole tail), it still forms a correctly patterned, functional organ, manages to get its data to the brain (*via* spinal cord connection) to enable vision (**Figure 6E**), and (if somewhere in the head) moves to the correct place during metamorphosis (Vandenberg et al., 2012). When cells are artificially made to be very large and have several times the normal genetic material, morphogenesis adapts to this and still builds an overall correct animal (Fankhauser, 1945a,b). These are goal-directed (in the cybernetic sense) processes because the system can reach a specific target morphology (and functionality) state despite perturbations or changes in local/starting conditions or the basic underlying components. Regeneration is the most familiar example of this, but is just a special case of the broader phenomenon of anatomical homeostasis. Homeostatic loops operating over large-scale anatomical states have several (closely related) key implications for the power and speed of evolution.

First, it greatly smoothes the fitness landscape. Consider two types of organisms: one whose subsystems mechanically follow a hardwired (genetically-specified) set of steps (A, passive, or merely structural modularity), and one whose modules optimize a reward function (B, multi-scale competency of modules). Mutations that would be detrimental in A (e.g., because they move the eye out of its optimal position) are neutral in B, because the competency of the morphogenetic subsystems repositions the eye even if it starts out somewhere else. Thus, MCA shields from selection some aspects of mutations' negative effects (which inevitably are the bulk of random mutations' consequences). The primary reason for the anatomical homeostasis seen in regulative development and regeneration may be for dealing, not with damage, but with deviations from target morphology induced by mutations. This is certainly true at the scale of tissues during the lifetime of an individual [as in the inverse relationship between regeneration and cancerous defection from large-scale target morphology (Levin, 2021b)], but may be true on evolutionary time scales as well.

Second, MCA reduces apparent pleiotropy—the fact that most mutations have multiple effects (Boyle et al., 2017). For example, a change in an important canonical signaling pathway such as Wnt or BMP (Raible and Ragland, 2005) is going to have numerous consequences for an organism. Suppose a mutation appears that moves the mouth off of its optimal position (bad for fitness) but also has some positive (adaptive) feature elsewhere in the body. In creatures of type A, the positive aspects of that mutation would never be seen by selection because the malfunctioning mouth would reduce the overall fitness or kill the individual outright. However, in creatures of type B, the mouth could move to its optimal spot (Vandenberg et al., 2012), enabling selection to independently evaluate the other consequence of that mutation. Creatures possessing MCA could reap the benefit of positive consequences of a mutation while masking its other effects *via* local adjustments to new changes that reduce the penalties (an important kind of buffering). In effect, evolution doesn't have to solve the very difficult search problem of "how to improve feature X without touching features Y. Z which already work well," and reaps massive efficiency (time) savings by not having to wait until the search process stumbles onto a way to directly encode an improvement that is either isolated from other features, or improves them all simultaneously (Wagner et al., 2007; Melo et al., 2016).

Third, MCA allows systems not only to solve problems, but also to exploit opportunities. A lineage has the chance to find out what pro-adaptive things a mutation can do, because competency hides the negative consequences. This gives time for new mutations to appear that hardwire the compensatory changes that had to be applied—an analogy to the proposed Baldwin effect (Hogenson, 2001; Downing, 2004; Robinson and Barron, 2017). This enables the opportunity to exploit the possibility space more freely, providing a kind of patience or reduction of the constraint that evolutionary benefits have to be immediate in order to propagate—it effectively reduces the short-sightedness of the evolutionary process. Indeed, multiscale competency is beneficial not only for natural evolution, but also for soft robotics and synthetic bioengineering because it helps cross the sim-to-real gap: models do not have to be 100% realistic to be predictively useful if the component modules can adaptively make up for some degree of deficiency in the controller design (Brooks, 1986).

Fourth, the homeostatic setpoint-seeking architecture makes the relationship between genotype and anatomical phenotype more linear (Muller and Schuppert, 2011; Lobo et al., 2014), improving controllability (Liu et al., 2011; Gao et al., 2014; Posfai et al., 2016). By using a top-down control layer to encode the patterns to which competent subunits operate, living systems do not need to solve the difficult inverse problems of what signals to send their subsystems to achieve high-level outcomes. Bioelectric pattern memories (such as the voltage distribution that tells wild-type planarian cells whether to build 1 head or 2) exploit a separation of data from the machine itself, which makes it much easier to make changes. Evolution does not need to find changes at the micro level but can also simply change the information encoded in the setpoints, such as the electric face prepattern (Vandenberg et al., 2012), which allows it to re-use the same exact implementation machinery to build something that can be quite different. The ability to rely on a non-zero IQ for your component modules (thus delegating and offloading complex regulatory chains) is an important affordance (Watson et al., 2010; Friston et al., 2012) for the evolutionary process. It means that the larger system's evolution is in effect searching an easier, less convoluted control, signaling, or reward space—this massive dimensionality reduction offers the same advantages human engineers have with agents on the right side of the persuadability scale. It is no accident that learning in the brain, and behavioral systems, eventually exapted this same architecture and indeed the exact same bioelectrical machinery to speed up the benefits of evolution.

A significant brake on the efficiency of evolution, as on machine learning (indeed, all learning) is credit assignment: which change or action led to the improvement or reward? When a collection of cells known as a "rat" learns to press a level and get a reward, no individual cell has the experience of interacting with a lever and receiving the nutrient. What enables the associative memory in this collective intelligence are the delay lines (nervous system) between the paws and the intestinal lining which provide a kind of patience—a tolerance of the temporal delay between the action and the reward and the ability to link extremely diverse modules on both ends (different kinds of actions can be linked to arbitrary rewards). MCA does the same thing for evolutionary learning (Watson et al., 2014, 2016; Power et al., 2015; Watson and Szathmary, 2016; Kouvaris et al., 2017), making it easier for systems to reap selection rewards for arbitrary moves in genotype space. This effectively raises the IQ of the evolutionary search process. Much as (**Figure 5**) an agent's sophistication can be gauged by how expertly and efficiently it navigates an arbitrary search space and its local optima, the traversal of the evolutionary search process can be made less short-sided by homeostatic activity within the physiological layer that sits between genotype and phenotype.

There is an adaptation tradeoff between robustness (e.g., morphogenesis to the same pattern despite interventions and changing conditions) and responsiveness to environment (context sensitivity), perhaps similar to the notion of criticality (Beggs, 2008; Hankey, 2015). The plasticity and goal-directedness of modules (as opposed to hardwired patterns) serve to reduce the sim-to-real gap (Kriegman et al., 2020b): because the current environment always offers novel challenges compared to prior experiences which evolution (or human design) uses to prepare responses, the MCA architecture doesn't take history too seriously, relying on plasticity and problem-solving more than on fine-tuning micromodels of what to do in specific cases. Biology reaps the benefits of both types of strategies by implementing anatomical homeostasis that coarse-grains robustness by making stability applying to large outcomes, such as overall anatomy, not to the microdetails of cell states. The scaling of homeostatic loops makes it possible to achieve both: consistent results and environmental sensitivity. These dynamics apply in various degrees to the numerous nested, adjacent, and overlapping sub-agents that make up any biological system. Cooperation results not from altruistic actions between Selves, but by the expansion of the borders of a single Self *via* scaling of the homeostatic loops. On this view, cancer cells are not more selfish than tissues—they are all equally selfish, but maintain goals appropriate to smaller scales of Selves. Indeed, even the parts of one normal body don't perfectly cooperate—this is as true in development (Gawne et al., 2020) as it is in cognitive science (Dorahy et al., 2014; Reinders et al., 2018, 2019). A picture is emerging of how evolution exploits the local competency of modules, competing and cooperating, to scale these subsystems' sensing, actuation, and setpoint memories to give rise to coherent larger-scale Selves. Overall, the TAME framework addresses functional aspects only, and is compatible with several views on phenomenal consciousness in compound Selves (Chalmers, 1996). However, it does have a few implications for the study of Consciousness.

CONSCIOUSNESS

While the TAME framework focuses on 3rd-person observable properties, it does make some commitments to ways of thinking about consciousness. Provisionally, I suggest that consciousness also comes in degrees and kinds (is not binary) for the same reasons argued for continuity of cognition: if consciousness is fundamentally embodied, the plasticity and gradual malleability of bodies suggests that it is a strong requirement for proponents of phase transitions to specify what kind of "atomic" (not further divisible) bodily change makes for a qualitative shift in capacity consciousness. Another implication of TAME is that while "embodiment" is critical for consciousness, it is not restricted to physical bodies acting in 3D space, but also includes perception-action systems working in all sorts of spaces. This implies, counter to many people's intuitions, that systems that operate in morphogenetic, transcriptional, and other spaces should also have some (if very minimal) degree of consciousness. This in turn suggests that an agent, such as a typical modern human, is really a patchwork of many diverse consciousnesses, only one of which is usually capable of verbally reporting its states (and, not surprisingly, given its limited access and self-boundary, believes itself to be a unitary, sole owner of the body).

What is necessary for consciousness? TAME's perspective is fundamentally that of the primacy of goal-directed activity. Thus, consciousness accompanies specific types of cognitive processes which exert energy toward goals, but as described

above, those processes can take forms very divergent from our typical brain-centered view. Unlike other panpsychist views, TAME does not claim that mind is inevitably baked in regardless of physical implementation or structure. Causal structure and cybernetic properties of the embodiment are key determinants of consciousness capacity. However, as the minimal degree of internal self-determination and goal-directedness is apparently present even in particles (Feynman, 1942; Georgiev and Georgiev, 2002; Ogborn et al., 2006; Kaila and Annila, 2008; Ramstead et al., 2019; Kuchling et al., 2020a), there may be no true "0" on the scale of consciousness in the universe. While simple accretion does not magnify the nano-goal-directed activity and indeterminate action of particles (e.g., rocks are not more conscious, and probably less, than particles in specific contexts), biological organization does amplify it, resulting in scaling up of sentience.

Of course, these implications will be unpalatable conclusions for many. It should be kept in mind that TAME is compatible with several different views on consciousness, and does not need to commit to one specific philosophy. It is fundamentally a framework for enabling empirical experiments, and its practical utility remains, regardless of the above speculations. Moreover, I remain skeptical about being able to say anything definitive about consciousness *per se* (as distinct from correlates of consciousness) from a 3rd-person, objective perspective. Thus, however unappealing the above view may be, I see no way of rigorously showing why any other claim about consciousness and what it requires is demonstrably better.

An emphasis on somatic plasticity has additional practical implications, being focused on the functional splitting and joining of agents' parts. For example, the ancient question of "where does it all come together?" in the brain, with respect to the unified character of consciousness, is one of those pseudo-problems that is dispelled by a framework like TAME that focuses on multi-scale architecture. How big should a place where it all comes together be? If it can be ∼140 mm wide, then the answer is, the whole brain. One could decide that it should be smaller (the human pineal gland is ∼7 mm wide), but then the question is, why not smaller still—given the cellular components of the pineal (or any piece of the brain) and the molecular organelles inside a pineal gland cell, one would always need to ask "but where does it all come together *inside there*?" of whatever piece of the brain is taken to be the seat of consciousness. The multi-scale nature of biology means that there is no privileged size scale for any homunculus.

Another important idea with respect to consciousness is "What is it like to be" a given agent (Nagel, 1974). Sensory augmentation, neural link technologies, and bioengineering produce tractable model systems in novel cognitive architectures, such as 2-headed planaria where the brains are connected by a central nervous system (**Figure 7B**), to help study the functional aspects of this cognitive re-shuffling. TAME's focus on the fact that all cognitive architectures are inevitably composites emphasizes that the parts can be rearranged; thus, the Subject of cognition can change "on the fly," not merely during evolutionary timescales. Thus, the basic question of philosophy of mind—what's it like to be animal X (Nagel, 1974)—is just a first-order step on a much longer journey. The second-order question is, what's it like to be a caterpillar, slowly *changing* into a butterfly as its brain is largely dissolved and reassembled into a different architecture for an animal whose sense organs, effectors, and perhaps overall Umwelt is completely different. All of this raises fascinating issues of first person experience not only in purely biological metamorphoses (such as human patients undergoing stem cell implants into their brains), but also technological hybrids such as brains instrumentized with novel sensory arrays, robotic bodies, software information systems, or brains functionally linked to other brains (Warwick et al., 1998; Demarse et al., 2001; Potter et al., 2003; Bakkum et al., 2007a,b; Tsuda et al., 2009; Cohen-Karni et al., 2012; Giselbrecht et al., 2013; Aaser et al., 2017; Ricotti et al., 2017; Ding et al., 2018; Mehrali et al., 2018; Anderson et al., 2020; Ando and Kanzaki, 2020; Merritt et al., 2020; Orive et al., 2020; Saha et al., 2020; Dong et al., 2021; Li et al., 2021; Pio-Lopez, 2021). The developmental approach to the emergence of consciousness on short, ontogenetic timescales complements the related question on phylogenetic timescales, and is likely to be a key component of mature theories in this field.

Most surprisingly, the plasticity and capacity for bioengineering and chimerization (recombination of biological and engineered parts in novel configurations) erases the sharp divide between first person and third person perspectives. This has been a fundamental, discrete distinction ever since Descartes, but the capacity for understanding and creating new combinations shows a continuum even in this basic distinction (**Figure 11**). The fact that Selves are not monadic means we can share parts with our subject of inquiry. If one has to *be* a system in order to truly know what it's like to be that system (1st person perspective), this is now possible, to various degrees, by physically merging one's cognitive architecture with that of another system. Of course, by strongly coupling to another agent, one doesn't remain the same and experience the other's consciousness; instead, a new Self is created that is a composite of the two prior individuals and has composite cognition. This is why consciousness research is distinct in strong degree from other scientific topics. One can observe gauges and instruments for 3rd-person science and remain the same Self (largely; the results of the observation may introduce small alterations in the cognitive structure). However, data on 1st person experiential consciousness cannot be taken in without fundamentally changing the Self (being an effective homunculus by watching the neuroscience data corresponding to the movies inside the heads of other people is impossible for the same reason that there is no homunculus in each of our heads). The study of consciousness, whether done *via* scientific tools or *via* the mind's own capacity to change itself, inevitably alters the Subject. Thus, standard (3rd-person) investigations of this process leave open the ancient question as to whether specific upgrades to cognition induce truly discontinuous jumps in consciousness. The TAME framework is not incompatible with novel discoveries about sharp phase transitions, but it takes the null hypothesis to be continuity, and it remains to be seen whether contrary evidence for truly sharp upgrades in consciousness can be provided. Future, radical brain-computer interfaces in human patients are perhaps one avenue where a subject undergoing such a change

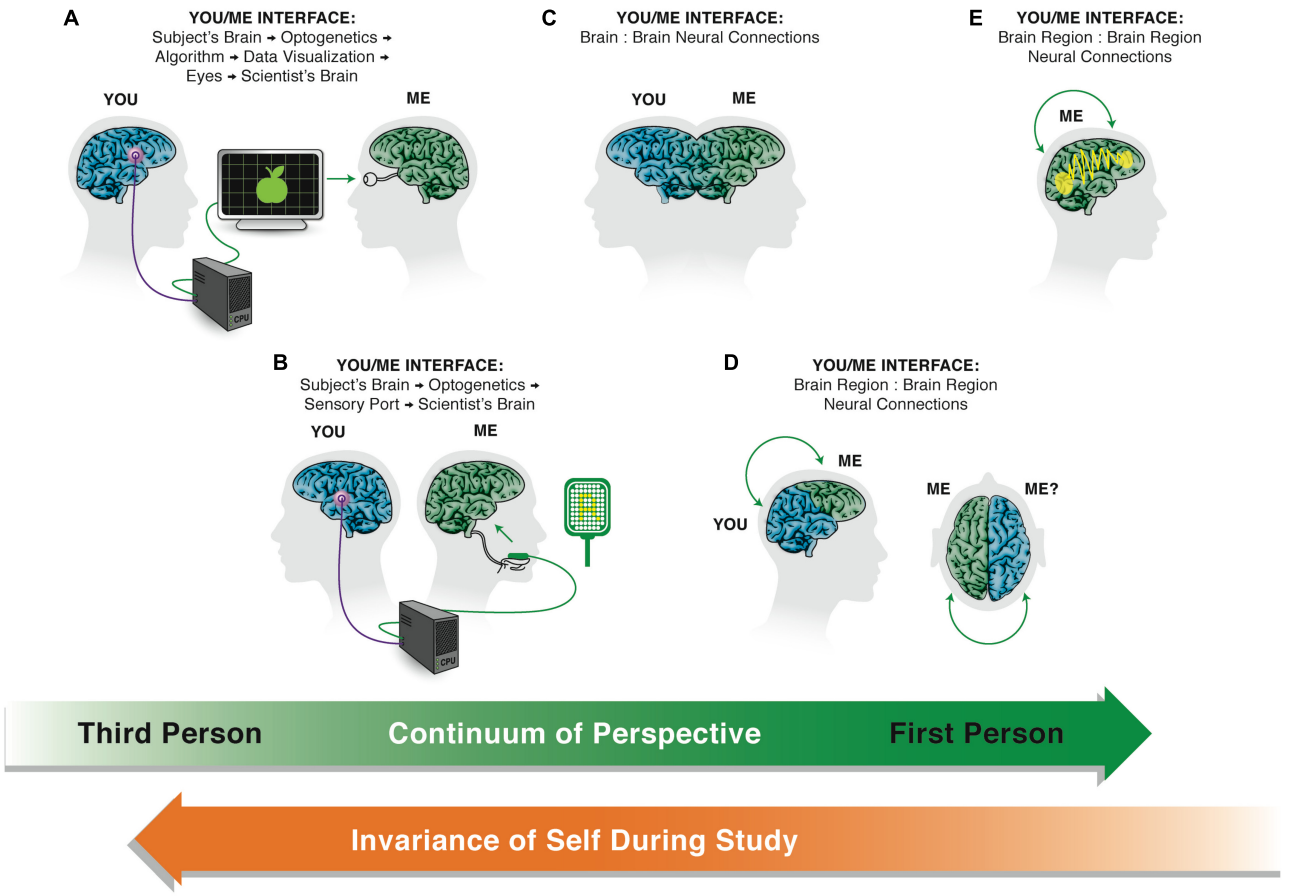


FIGURE 11 | Technology reveals gradualism in Descartes' cut. The apparent fundamental gulf between first person perspective (what is it like to *be* a specific Self) and third person perspective (external scientific study of that cognitive system) can be seen to also be a gradual continuum, when modern technology is used to expand heterophenomenology (Dennett, 1991). On the left side of the continuum **(A)** is a traditional 3-rd person scenario of an agent studying another by measuring its physical states: cognitive states can be inferred but not directly experienced (correlates of consciousness), giving rise to the problem of other minds, and a firm distinction between "you" and "me." However, sensory substitution and augmentation technology now enables the plugging of various peripherals and sensors directly into the nervous system of subjects, and thus it is possible to connect the output of electrophysiology equipment directly into the subject's brain **(B)**. For example, if the output of a multielectrode array recording neural activity of subject #1 is connected directly to the brainport device (Danilov and Tyler, 2005) of subject #2 (e.g., a scientist), this allows #1's mental states to more directly provide input into #2's sensory stream. This can be made even more direct by fusing portions of two brains directly, during embryogenesis, illustrating that the strength of boundaries between "you" and "me" is variable; this configuration may seem far-fetched, but note that it can be readily produced today in animal model systems, and the only barrier to such configurations is ethical, not empirical or logical **(C)**. It's critical to note that these fusion experiments are not just aberrant corner cases, because all brains are *already* fusions of neural modules. Single subjects' brains consist of two hemispheres which must communicate to give rise to a coherent, centralized perception of "me" despite being made of communicating parts, and can be dissociated by commissurotomy **(D)**. Indeed, beyond the two hemispheres, any brain is a collective of smaller active subunits **(E)** that must all communicate as a collection of cells (each neuron is part of the neighboring neural cell's "external environment"). This gradient of diverse connections, whether electronic or biological, between and within brains and brain components can be reproduced or expanded upon to whatever degree necessary by biological or technological fusion among subjects. The technological aspect of TAME is that we must develop frameworks that deal not only with standard embodiments of mind as happened to be produced by the path evolution took through life on Earth, but all logically and empirically possible configurations that could evolve, be designed, or both, whether on Earth or in exobiological contexts. The hierarchical, not monadic, structure of cognitive substrates means that the relationship between the parts of one Self and that between a Self and an object of external study is a continuum, not a discrete natural kind. This suggests a key way that actual Consciousness can be studied—by becoming inherently a participant in the experiment, so as to study it from a first-person perspective. Importantly however, what happens when one fuses cognitive systems with their subject of study is that a new Self appears (a composite cognitive system), showing that the Self can remain invariant while pursuing scientific study of functional cognition and behavior (the left of the spectrum), but essentially must change in order to gain first-hand knowledge of consciousness in other cognitive systems. Images are courtesy of Jeremy Guay of Peregrine Creative.

can convince themselves, and perhaps others, that a qualitative, not continuous, change in their consciousness had occurred.

With respect to the question of *consciousness per se*, as opposed to neural or behavioral correlates of consciousness, we have one major functional tool: general anesthesia. It is remarkable that we can readily induce a state in which all the individual cells are fine and healthy, but the larger Self is simply gone [although, some of the parts can continue to learn during this time (Ghoneim and Block, 1997)]. Interestingly, general anesthetics are gap junction blockers (Wentlandt et al., 2006): consistent with the cognitive

scaling example above, shutting down electrical communication among the cells leads to a disappearance of the higher-level computational layer while the cellular network is disrupted. GJ blockers are used to anesthetize living beings ranging across plants, Hydra, and human subjects (Gremiaux et al., 2014). It is amazing that the same Self (with memories and other properties) returns, when the anesthetic is removed. Of course, the Self does not return immediately, as shown by the many hallucinatory (Saniova et al., 2009; Kelz et al., 2019) experiences of people coming out of general anesthesia—it takes some time for the brain to return to the correct global bioelectric state once the network connections are allowed again (meta-stability) (Rabinovich et al., 2008). Interestingly, and in line with the proposed isomorphism between cognition and morphogenesis, gap junction blockade has exactly this effect in regeneration: planaria briefly treated with GJ blocker regenerate heads of other species, but eventually snap out of it and remodel back to their correct target morphology (Emmons-Bell et al., 2015). It is no accident that the same reagents cause drastic changes in the high-level Selves in both behavioral and morphogenetic contexts: evolution uses the same scheme (GJ-mediated bioelectrical networks) to implement both.

The epistemic problem of Other Minds has been framed to imply that we cannot directly ever be sure how much or what kind of consciousness exists in any particular system under study. The TAME framework reminds us that this is true even for components of ourselves (like the non-verbal brain hemisphere). Perhaps the confabulation system enables one part of our mind to estimate the agency of other parts (the feelings of consciousness and free will) and develop models useful for prediction and control, applying in effect the empirical criteria for persuadability internally. The ability to develop a "theory of mind" about external agents can readily be turned inward, in a composite Self.

Are all cognitive systems conscious? The TAME framework is compatible with several views on the nature of consciousness. However, the evolutionary conservation of mechanisms between brains and their non-neural precursors has an important consequence for the question of where consciousness could be found. To the extent that one believes that mechanisms in the brain enable consciousness, all of the same machinery and many similar functional aspects are found in many other places in the body and in other constructs. TAME emphasizes that there is no principled way to restrict consciousness to "human-like, full-blown sophisticated brains," which means one has to seriously consider degrees of consciousness in other organs, tissues, and synthetic constructs that have the same features neurons and their networks do (Trewavas and Baluska, 2011; Baluska et al., 2016, 2021; Baluska and Reber, 2019). The fundamental gradualism of this framework suggests that whatever consciousness is, some variant and degree thereof has to be present very widely across autopoietic systems. TAME is definitely incompatible with binary views that cut off consciousness at a particular sharp line and it suggests no obvious way to define cognitive systems that have no consciousness whatsoever. A big open question is whether the continuum of cognition (and consciousness) contains a true "0" or only infinitesimal levels for very modest agents. One is tempted to imagine what properties a *truly minimal* agent would have to have; not being fully constrained by local forces, and ability to pursue goals, both seem key, and both of these are present to a degree in even single particles (*via* quantum indeterminacy and least action behavior). The type and degree of scaling (or lack thereof) of these capacities in bulk inorganic matter vs. highly-organized living forms is a fertile area for future development of TAME and will be explored in forthcoming work.

CONCLUSION

A More Inclusive Framework for Cognition

Regenerating, physiological, and behaving systems use effort (energy) to achieve defined, adaptive outcomes despite novel circumstances and unpredictable perturbations. That is a key invariant for cognition; differences in substrate, scale, or origin story among living systems are not fundamental, and obscure an important way to unify key properties of life: the ability to deploy intelligence for problem-solving in diverse domains. Modern theories of Mind must eventually handle the entire option space for intelligent agents, which not only contains the familiar advanced animals we see on Earth, but can also subsume ones consisting of radically different materials, ones created by synthetic bioengineering or combinations of evolution and rational design in the lab, and ones of exobiological as well as possible terrestrial origins. The advances of engineering confirm and put into practice an idea that was already entailed by evolution: that cognitive traits, like all other traits, evolved from humbler variants, forming a continuum. There are no biologically-valid binary categories in this space. Take the prevalent legal definition of human "adults," who snap into being at the age of 18; such binary views on cognitive properties are fictitious coarse-grainings useful for our legal system to operate, but no more than that. There is no bright line between "truly cognitive" and "pseudo cognitive" that can ever be drawn between two successive members of an evolutionary lineage. The error of "committing Anthropomorphism" is a pseudo-scientific "folk" notion useful for only the most trivial examples of failure to scale down complex claims proportionally to simpler systems; engineering requires us to determine what level of cognitive model enables the most fruitful prediction and control.

Every intelligence is a collective intelligence, and the modular, multi-scale architecture of life means that we are a holobiont in more than just the sense of having a microbiome (Chiu and Gilbert, 2015)—we are all patchworks of overlapping, nested, competing, and cooperating agents that have homeostatic (goal-directed) activity within their self-constructed virtual space at a scale that determines their cognitive sophistication. A highly tractable model system for unconventional cognition, in which these processes and the scaling of Selves can not only be seen but can also be manipulated, is morphogenetic homeostasis. The process of construction and remodeling (toward anatomical features) of cellular collectives shows crucial isomorphism to cognitive aspects of the many-into-one binding like credit assignment, learning, stress reduction, etc. The partial wiping

of ownership information on permanent signals makes gap junctional coupling an excellent minimal model system for thinking about biological mechanisms that scale cognition while enabling co-existence of subunits with local goals (multiple levels of overlapping Selves, whose scale and borders are porous and can change during the lifetime of the agent). However, many other substrates can no doubt fulfill the same functions.

Next Steps: Conceptual and Empirical Research Programs

The TAME framework is conceptually incomplete in important ways. On-going development is proceeding along lines including merging with other frameworks such as Active Inference (Friston, 2013; Badcock et al., 2019; Ramstead et al., 2019), Rosen's (M,R) and Anticipatory Systems (Rosen, 1973, 1979, 1985; Nasuto and Hayashi, 2016), and recent advances in information theory as applied to individuality and scaling of causal power (Hoel et al., 2013, 2016; Krakauer et al., 2014; Daniels et al., 2016). It will be critical to more rigorously develop the waypoints along the Persuadability Continuum, including understanding of what an "increased capacity" human (or non-human) would be like, in contrast to the "diminished capacity" with which we are well familiar from legal proceedings [the right side of the continuum, corresponding to radically expanded cognitive light cones (Śāntideva Bstan 'dzin rgya m and Comité de traduction Padmakara, 2006)].

The TAME framework suggests numerous practical research directions immediately within reach (some of which are already pursued in our group), including developing biomedically-actionable models of morphogenetic plasticity and robustness as meta-cognitive error correction mechanisms, tissue training paradigms for anatomical and physiological outcomes, exploiting learning properties of pathway models for regenerative medicine (Herrera-Delgado et al., 2018; Biswas et al., 2021), and creation of AI platforms based on multi-scale agency architectures that do not rely on neuromorphic principles.

Beyond Basic Science: Up-to-Date Ethics

The TAME framework also has implications for ethics in several ways. The current emphasis for ethics is on whether bioengineered constructs (e.g., neural cell organoids) are sufficiently like a human brain or not (Hyun et al., 2020), as a criterion for acceptability. Likewise, existing efforts to extend ethics focus on natural, conventional evolutionary products such as invertebrates (Mikhalevich and Powell, 2020). TAME suggests that this is insufficient, because many different architectures for cognition are possible (and will be realized)—similarity to human brains is too parochial and limiting a marker for entities deserving of protection and other moral considerations. We must develop a new ethics that recognizes the diversity of possible minds and bodies, especially since combinations of biological, engineered, and software systems are, and increasingly will be, developed. What something looks like and how it originated (Levin et al., 2020; Bongard and Levin, 2021) will no longer be a good guide when we are confronted with a myriad of creatures that cannot be comfortably placed within the familiar Earth's phylogenetic tree.

Bioengineering of novel Selves raises our moral responsibility. For eons, humans have been creating and releasing into the world advanced, autonomous intelligences—*via* pregnancy and birth of other humans. This, in Dennett's phrase, has been achieved until now *via* high levels of "competency without comprehension" (Dennett, 2017); however, we are now moving into a phase in which we create beings *via* comprehension—with rational control over their structure and cognitive capacities, which brings additional responsibility. A new ethical framework will have to be formed without reliance on binary folk notions such as "machine," "robot," "evolved," "designed," etc., because these categories are now seen to not be crisp natural kinds. Instead, wider approaches (such as Buddhist concern for all sentient beings) may be needed to act ethically with respect to agents that have preferences, goals, concerns, and cognitive capacity in very unfamiliar guises. TAME seeks to break through the biases around contingent properties that drive our estimates of who or what deserves proper treatment, to develop a rational, empirically-based mechanism for recognizing Selves around us.

Another aspect of ethics is the discussion of limits on technology. Much of it is often driven by a mindset of making sure we don't run afoul of the risks of negative uses of specific technologies (e.g., genetically-modified organisms in ecosystems). This is of course critical with respect to the new bioengineering capabilities. However, such discussions often are one-sided, framed as if the *status quo* was excellent, and our main goal is simply to not make things worse. This is a fundamental error which neglects the opportunity cost of failing to fully exploit the technologies which could drive advances in the control of biology. The *status quo* is not perfect—society faces numerous problems including disparities of quality of life across the globe, incredible suffering from unsolved medical needs, climate change, etc. It must be kept in mind that along with the need to limit negative consequences of scientific research, there is a *moral imperative* to advance aspects of research programs that will (for example) enable the cracking of the morphogenetic code to revolutionize regenerative medicine far beyond what genomic editing and stem cell biology can do alone (Levin, 2011).

The focus on risk arises from a feeling that we should not "mess with nature," as if the existing structures (from anatomical order to ecosystems) are ideal, and our fumbling attempts will disrupt their delicate balance. While being very careful with powerful advances, it must also be kept in mind that existing balance (i.e., the homeostatic goals of systems from cells to species in the food web) was not achieved by optimizing happiness or any other quality commensurate with modern values: it is the result of dynamical systems properties shaped by the frozen accidents of the meanderings of the evolutionary process and the harsh process of selection for survival capacity. We have the opportunity to use rational design to do better than the basic mechanisms of evolution allow.

Importantly, current technologies are forcing us to confront an existential risk. Swarm robotics, Internet of Things, AI, and similar engineering efforts are going to be creating numerous complex, goal-driven systems made up of competent parts. We

currently have no mature science of where the goals of such novel Selves come from. TAME reminds us that it is essential to understand how goals of composite entities arise and how they can be predicted and controlled. To avoid the Skynet scenario (Bostrom, 2015), it is imperative to study the scaling of cognition in diverse substrates, so that we can ensure that the goals of powerful, distributed novel beings align with ours.

Given the ability of human subunits to merge into even larger (social) structures, how do we construct higher-order Selves that promote flourishing for all? The multicellularity-cancer dynamic (**Figure 9**) suggests that tight functional connections that blur cognitive boundaries among subunits is a way to increase cooperation and cognitive capacity. However, simply maximizing loss of identity into massive collectivism is a well-known failure at the social level, always resulting in the same dynamic: the goals of the whole diverge sharply from those of the parts, which become as disposable to the larger social Self as shed skin cells are to us. Thus, the goal of this research program beyond biology is the search for optimal binding policies between subunits, which optimize the tradeoffs needed to maximize individual goals and well-being (preserving freedom or empowerment) while reaping the benefits of a scaled-up Self at the level of groups and entire societies. While the specific binding mechanisms used by evolution are not guaranteed to be the policies we want at the social level, the study of these are critical for jump-starting a rigorous program of research into possible ways of scaling that could have social relevance. These issues have been previously addressed in the context of evolutionary dynamics and game theory (Maynard Smith and Szathmáry, 1995; Michod and Nedelcu, 2003; Van Baalen, 2013), but can be significantly expanded using the TAME framework.

In the end, important ethical questions around novel agents made of combinations of hardware, software, evolved, and designed components always come back to the nature of the Self. The coherence of a mind, along with its ability to pursue goal-directed activity, is central to our notions of moral responsibility in the legal sense: diminished capacity, and soon, enhanced capacity, to make *choices* is a pillar for social structures. Mechanist views of cause and effect in the neuroscience of behavior have been said to erode these classical notions. Rather than reduce Selves (to 0, in some eliminativist approaches), TAME (Levin, 2022) finds novel Selves all around us. We see more agency, not less, when evolution and cell biology are taken seriously (Levin and Dennett, 2020). The cognitive Self is not an illusion; what is an illusion is that there is only one, permanent, privileged Self that has to arise entirely bottom-up through the hill-climbing process of evolution. Our goal, at the biomedical, personal, and social levels should not be to destroy or minimize the Self but to recognize it in all its guises, understand its transitions, and enlarge its cognitive capacity toward the well-being of other Selves.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

ML developed all the ideas and wrote the entire manuscript.

FUNDING

I gratefully acknowledge support by the Paul G. Allen Frontiers Group (*via* an Allen Discovery Center Award 12171), the Templeton World Charity Foundation (WCF0089/AB55 and TWCF0140), and John Templeton Fund (Grant 62212 from the John Templeton Foundation). The funders had no role in determining the content of this manuscript.

ACKNOWLEDGMENTS

I would like to thank Dora Biro, Joshua Bongard, Avery Caulfield, Anna Ciaunica, Pranab Das, Daniel Dennett, Thomas Doctor, Bill Duane, Christopher Fields, Adam Goldstein, EJ, Aysja Johnson, Jeantine Lunshof, Santosh Manicka, Patrick McMillen, Aniruddh Patel, Giovanni Pezzulo, Andrew Reynolds, Elizaveta Solomonova, Matthew Simms, Richard Watson, Olaf Witkowski, Rafael Yuste, and numerous others from the Levin Lab and the Diverse Intelligences community for helpful conversations and discussions, as well as comments on versions of this manuscript. I would also like to thank the three reviewers of the manuscript for important critiques that led to improvement. This manuscript is dedicated to my mother, Luba Levin, who while not having been a scientist, always modeled a deep understanding of, and care for, the multi-scale agency abundant in the world.

REFERENCES

Aaser, P., Knudsen, M., Ramstad, O. H., van de Wijdeven, R., Nichele, S., Sandvig, I., et al. (2017). "Towards making a cyborg: a closed-loop reservoir-neuro system," in *ECAL 2017: The 14th European Conference on Artificial Life* (Lyon: MIT Press), 430–437.

Abraham, W. C., Jones, O. D., and Glanzman, D. L. (2019). Is plasticity of synapses the mechanism of long-term memory storage? *NPJ Sci. Learn.* 4:9. doi: 10.1038/s41539-019-0048-y

Adams, D. S., Lemire, J. M., Kramer, R. H., and Levin, M. (2014). Optogenetics in developmental biology: using light to control ion flux-dependent signals in Xenopus embryos. *Int. J. Dev. Biol.* 58, 851–861. doi: 10.1387/ijdb.140207ml

Adams, D. S., Masi, A., and Levin, M. H. (2007). H+ pump-dependent changes in membrane voltage are an early mechanism necessary and sufficient to induce Xenopus tail regeneration. *Development* 134, 1323–1335. doi: 10.1242/dev.02812

Adams, D. S., Tseng, A. S., and Levin, M. (2013). Light-activation of the Archaerhodopsin H(+)-pump reverses age-dependent loss of vertebrate regeneration: sparking system-level controls in vivo. *Biol. Open* 2, 306–313. doi: 10.1242/bio.20133665

Adams, D. S., Uzel, S. G., Akagi, J., Wlodkowic, D., Andreeva, V., Yelick, P. C., et al. (2016). Bioelectric signalling via potassium channels: a mechanism for craniofacial dysmorphogenesis in KCNJ2-associated Andersen-Tawil Syndrome. *J. Physiol.* 594, 3245–3270. doi: 10.1113/JP271930

Alloway, T. M. (1972). Retention of learning through metamorphosis in grain beetle (Tenebrio-Molitor). *Am. Zool.* 12, 471–472.

Ameriks, K. (1976). Personal identity and memory transfer. *Southern J. Phil.* 14, 385–391. doi: 10.1111/j.2041-6962.1976.tb01295.x

Anderson, M. J., Sullivan, J. G., Horiuchi, T., Fuller, S. B., and Daniel, T. L. (2020). A bio-hybrid odor-guided autonomous palm-sized air vehicle. *Bioinspir. Biomim.* 16:026002. doi: 10.1088/1748-3190/abbd81

Ando, N., and Kanzaki, R. (2020). Insect-machine hybrid robot. *Curr. Opin. Insect. Sci.* 42, 61–69. doi: 10.1016/j.cois.2020.09.006

Ariazi, J., Benowitz, A., De Biasi, V., Den Boer, M. L., Cherqui, S., Cui, H., et al. (2017). Tunneling nanotubes and gap junctions-their role in long-range intercellular communication during development, health, and disease conditions. *Front. Mol. Neurosci.* 10:333. doi: 10.3389/fnmol.2017.00333

Armstrong, J. D., de Belle, J. S., Wang, Z., and Kaiser, K. (1998). Metamorphosis of the mushroom bodies; large-scale rearrangements of the neural substrates for associative learning and memory in Drosophila. *Learn. Mem.* 5, 102–114.

Auletta, G. (2011). Teleonomy: the feedback circuit involving information and thermodynamic processes. *J. Mod. Phys.* 2, 136–145. doi: 10.4236/jmp.2011.23021

Bach-y-Rita, P. (1981). Brain plasticity as a basis of the development of rehabilitation procedures for hemiplegia. *Scand. J. Rehabil. Med.* 13, 73–83.

Bach-y-Rita, P., Collins, C. C., Saunders, F. A., White, B., and Scadden, L. (1969). Vision substitution by tactile image projection. *Nature* 221, 963–964. doi: 10.1038/221963a0

Badcock, P. B., Friston, K. J., and Ramstead, M. J. D. (2019). The hierarchically mechanistic mind: a free-energy formulation of the human psyche. *Phys. Life Rev.* 31:104–121. doi: 10.1016/j.plrev.2018.10.002

Bakkum, D. J., Gamblen, P. M., Ben-Ary, G., Chao, Z. C., and Potter, S. M. (2007b). MEART: the semi-living artist. *Front. Neurorobot.* 1:5. doi: 10.3389/neuro.12.005.2007

Bakkum, D. J., Chao, Z. C., Gamblen, P., Ben-Ary, G., Shkolnik, A. G., DeMarse, T. B., et al. (2007a). "Embodying cultured networks with a robotic drawing arm," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Vol. 2007 (Lyon), 2996–2999. doi: 10.1109/IEMBS.2007.4352959.

Balazsi, G., van Oudenaarden, A., and Collins, J. J. (2011). Cellular decision making and biological noise: from microbes to mammals. *Cell* 144, 910–925. doi: 10.1016/j.cell.2011.01.030

Baluška, F., and Levin, M. (2016). On having no head: cognition throughout biological systems. *Front. Psychol.* 7:902. doi: 10.3389/fpsyg.2016.00902

Baluska, F., and Mancuso, S. (2012). Ion channels in plants: from bioelectricity, via signaling, to behavioral actions. *Plant Signal. Behav.* 8:e23009. doi: 10.4161/psb.23009

Baluska, F., and Reber, A. (2019). Sentience and consciousness in single cells: how the first minds emerged in unicellular species. *BioEssays* 41:e1800229. doi: 10.1002/bies.201800229

Baluska, F., Miller, W. B. Jr., and Reber, A. S. (2021). Biomolecular basis of cellular consciousness via subcellular nanobrains. *Int. J. Mol. Sci.* 22:2545. doi: 10.3390/ijms22052545

Baluska, F., Yokawa, K., Mancuso, S., and Baverstock, K. (2016). Understanding of anesthesia - Why consciousness is essential for life and not based on genes. *Commun. Integr. Biol.* 9:e1238118. doi: 10.1080/19420889.2016.1238118

Barilan, Y. M. (2003). One or two: an examination of the recent case of the conjoined twins from Malta. *J. Med. Philos.* 28, 27–44. doi: 10.1076/jmep.28.1.27.14176

Bates, E. (2015). Ion channels in development and cancer. *Annu. Rev. Cell Dev. Biol.* 31, 231–247. doi: 10.1146/annurev-cellbio-100814-125338

Batterman, R. (2015). "Autonomy and scales," in *Front Collection*, eds B. Falkenburg and M. Morrison (Berlin: Springer), 115–135. doi: 10.1007/978-3-662-43911-1_7

Batterman, R. W., and Rice, C. C. (2014). Minimal model explanations. *Philos. Sci.* 81, 349–376. doi: 10.1086/676677

Bayne, T., Brainard, D., Byrne, R. W., Chittka, L., Clayton, N., Heyes, C., et al. (2019). What is cognition? *Curr. Biol.* 29, R608–R615. doi: 10.1016/j.cub.2019.05.044

Bedecarrats, A., Chen, S., Pearce, K., Cai, D., and Glanzman, D. L. (2018). RNA from trained aplysia can induce an epigenetic engram for long-term sensitization in untrained aplysia. *eNeuro* 5:ENEURO.0038-18.2018. doi: 10.1523/ENEURO.0038-18.2018

Beekman, M., and Latty, T. (2015). Brainless but multi-headed: decision making by the acellular slime mould Physarum polycephalum. *J. Mol. Biol.* 427, 3734–3743. doi: 10.1016/j.jmb.2015.07.007

Beer, R. D. (2014). The cognitive domain of a glider in the game of life. *Artif. Life* 20, 183–206. doi: 10.1162/ARTL_a_00125

Beer, R. D. (2015). Characterizing autopoiesis in the game of life. *Artif. Life* 21, 1–19. doi: 10.1162/ARTL_a_00143

Beer, R. D., and Williams, P. L. (2015). Information processing and dynamics in minimally cognitive agents. *Cogn. Sci.* 39, 1–38. doi: 10.1111/cogs.12142

Beggs, J. M. (2008). The criticality hypothesis: how local cortical networks might optimize information processing. *Philos. Trans. A Math. Phys. Eng. Sci.* 366, 329–343. doi: 10.1098/rsta.2007.2092

Belwafi, K., Gannouni, S., and Aboalsamh, H. (2021). Embedded brain computer interface: state-of-the-art in research. *Sensors* 21:4293. doi: 10.3390/s21134293

Berdahl, A. M., Kao, A. B., Flack, A., Westley, P. A. H., Codling, E. A., Couzin, I. D., et al. (2018). Collective animal navigation and migratory culture: from theoretical models to empirical evidence. *Philos. Transac. R. Soc. B Biol. Sci.* 373:20170009. doi: 10.1098/rstb.2017.0009

Birch, J., Ginsburg, S., and Jablonka, E. (2020). Unlimited associative learning and the origins of consciousness: a primer and some predictions. *Biol. Philos.* 35:56. doi: 10.1007/s10539-020-09772-0

Bisping, R., Oehlert, U., Reinauer, H., and Longo, N. (1971). Negative and positive memory transfer through RNA in instrumentally conditioned goldfish. *Stud. Psychol.* 13, 181–190.

Biswas, S., Manicka, S., Hoel, E., and Levin, M. (2021). Gene regulatory networks exhibit several kinds of memory: quantification of memory in biological and random transcriptional networks. *iScience* 24:102131. doi: 10.1016/j.isci.2021.102131

Blackiston, D. J., and Levin, M. (2013). Ectopic eyes outside the head in Xenopus tadpoles provide sensory data for light-mediated learning. *J. Exp. Biol.* 216(Pt. 6), 1031–1040. doi: 10.1242/jeb.074963

Blackiston, D. J., Silva Casey, E., and Weiss, M. R. (2008). Retention of memory through metamorphosis: can a moth remember what it learned as a caterpillar? *PLoS One* 3:e1736. doi: 10.1371/journal.pone.0001736

Blackiston, D. J., Vien, K., and Levin, M. (2017). Serotonergic stimulation induces nerve growth and promotes visual learning via posterior eye grafts in a vertebrate model of induced sensory plasticity. *NPJ Regen. Med.* 2:8. doi: 10.1038/s41536-017-0012-5

Blackiston, D., Adams, D. S., Lemire, J. M., Lobikin, M., and Levin, M. (2011). Transmembrane potential of GlyCl-expressing instructor cells induces a neoplastic-like conversion of melanocytes via a serotonergic pathway. *Dis. Models Mech.* 4, 67–85. doi: 10.1242/dmm.005561

Blackiston, D., Lederer, E. K., Kriegman, S., Garnier, S., Bongard, J., and Levin, M. (2021). A cellular platform for the development of synthetic living machines. *Sci. Robot* 6:eabf1571. doi: 10.1126/scirobotics.abf1571

Blackiston, D., Shomrat, T., and Levin, M. (2015). The stability of memories during brain remodeling: a perspective. *Commun. Integr. Biol.* 8:e1073424. doi: 10.1080/19420889.2015.1073424

Bongard, J., and Levin, M. (2021). Living things are not (20th Century) machines: updating mechanism metaphors in light of the modern science of machine behavior. *Front. Ecol. Evol.* 9:650726. doi: 10.3389/fevo.2021.650726

Bongard, J., Zykov, V., and Lipson, H. (2006). Resilient machines through continuous self-modeling. *Science* 314, 1118–1121. doi: 10.1126/science.1133687

Bonzanni, M., Rouleau, N., Levin, M., and Kaplan, D. L. (2020). Optogenetically induced cellular habituation in non-neuronal cells. *PLoS One* 15:e0227230. doi: 10.1371/journal.pone.0227230

Bostrom, N. (2015). *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.

Boussard, A., Delescluse, J., Perez-Escudero, A., and Dussutour, A. (2019). Memory inception and preservation in slime moulds: the quest for a common

mechanism. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 374:20180368. doi: 10.1098/rstb.2018.0368

Boyle, E. A., Li, Y. I., and Pritchard, J. K. (2017). An expanded view of complex traits: from polygenic to omnigenic. *Cell* 169, 1177–1186. doi: 10.1016/j.cell.2017.05.038

Brooks, R. A. (1986). A robust layer control system for a mobile robot. *IEEE J. Robot. Automation* 2, 14–23.

Brugger, P., Macas, E., and Ihlemann, J. (2002). Do sperm cells remember? *Behav. Brain Res.* 136, 325–328. doi: 10.1016/s0166-4328(02)00127-4

Bubenik, A. (1966). *Das Geweih*. Hamburg: Paul Parey Verlag.

Bubenik, A. B., and Pavlansky, R. (1965). Trophic responses to trauma in growing antlers. *J. Exp. Zool.* 159, 289–302. doi: 10.1002/jez.1401590302

Bucher, D., and Anderson, P. A. V. (2015). Evolution of the first nervous systems – what can we surmise? *J. Exp. Biol.* 218, 501–503. doi: 10.1242/jeb.111799

Busse, S. M., McMillen, P. T., and Levin, M. (2018). Cross-limb communication during Xenopus hindlimb regenerative response: non-local bioelectric injury signals. *Development* 145:dev164210. doi: 10.1242/dev.164210

Buznikov, G. A., Peterson, R. E., Nikitina, L. A., Bezuglov, V. V., and Lauder, J. M. (2005). The pre-nervous serotonergic system of developing sea urchin embryos and larvae: pharmacologic and immunocytochemical evidence. *Neurochem. Res.* 30, 825–837. doi: 10.1007/s11064-005-6876-6

Calvo, P., and Friston, K. (2017). Predicting green: really radical (plant) predictive processing. *J. R. Soc. Interface* 14:20170096. doi: 10.1098/rsif.2017.0096

Camley, B. A. (2018). Collective gradient sensing and chemotaxis: modeling and recent developments. *J. Phys. Condens. Matter* 30:223001. doi: 10.1088/1361-648X/aabd9f

Cartmill, M. (2017). Convergent? Minds? Some questions about mental evolution. *Interface Focus* 7:20160125. doi: 10.1098/rsfs.2016.0125

Cebrià, F., Adell, T., and Saló, E. (2018). Rebuilding a planarian: from early signaling to final shape. *Int. J. Dev. Biol.* 62, 537–550. doi: 10.1387/ijdb.180042es

Cervera, J., Levin, M., and Mafe, S. (2020a). Bioelectrical coupling of single-cell states in multicellular systems. *J. Phys. Chem. Lett.* 3234–3241. doi: 10.1021/acs.jpclett.0c00641

Cervera, J., Meseguer, S., Levin, M., and Mafe, S. (2020b). Bioelectrical model of head-tail patterning based on cell ion channels and intercellular gap junctions. *Bioelectrochemistry* 132:107410. doi: 10.1016/j.bioelechem.2019.107410

Cervera, J., Pai, V. P., Levin, M., and Mafe, S. (2019b). From non-excitable single-cell to multicellular bioelectrical states supported by ion channels and gap junction proteins: Electrical potentials as distributed controllers. *Prog. Biophys. Mol. Biol.* 149, 39–53. doi: 10.1016/j.pbiomolbio.2019.06.004

Cervera, J., Manzanares, J. A., Mafe, S., and Levin, M. (2019a). Synchronization of bioelectric oscillations in networks of nonexcitable cells: from single-cell to multicellular states. *J. Phys. Chem. B* 123, 3924–3934. doi: 10.1021/acs.jpcb.9b01717

Cervera, J., Pietak, A., Levin, M., and Mafe, S. (2018). Bioelectrical coupling in multicellular domains regulated by gap junctions: a conceptual approach. *Bioelectrochemistry* 123, 45–61. doi: 10.1016/j.bioelechem.2018.04.013

Chalmers, D. (1996). *The Conscious Mind*. New York, NY: Oxford University Press.

Chalmers, D. (2013). Panpsychism and panprotopsychism. *Amherst Lecture Philosophy* 8.

Chamola, V., Vineet, A., Nayyar, A., and Hossain, E. (2020). Brain-computer interface-based humanoid control: a review. *Sensors* 20:3620. doi: 10.3390/s20133620

Chao, Z. C., Bakkum, D. J., and Potter, S. M. (2008). Shaping embodied neural networks for adaptive goal-directed behavior. *PLoS Comput. Biol.* 4:e1000042. doi: 10.1371/journal.pcbi.1000042

Chen, S., Cai, D., Pearce, K., Sun, P. Y., Roberts, A. C., and Glanzman, D. L. (2014). Reinstatement of long-term memory following erasure of its behavioral and synaptic expression in Aplysia. *Elife* 3:e03896. doi: 10.7554/eLife.03896

Chernet, B. T., Adams, D. S., Lobikin, M., and Levin, M. (2016). Use of genetically encoded, light-gated ion translocators to control tumorigenesis. *Oncotarget* 7, 19575–19588. doi: 10.18632/oncotarget.8036

Chernet, B. T., and Levin, M. (2013a). Endogenous voltage potentials and the microenvironment: bioelectric signals that reveal, induce and normalize cancer. *J. Clin. Exp. Oncol.* Suppl. 1, S1-002. doi: 10.4172/2324-9110.S1-002

Chernet, B. T., and Levin, M. (2013b). Transmembrane voltage potential is an essential cellular parameter for the detection and control of tumor development in a Xenopus model. *Dis. Models Mech.* 6, 595–607. doi: 10.1242/dmm.010835

Chernet, B. T., and Levin, M. (2014). Transmembrane voltage potential of somatic cells controls oncogene-mediated tumorigenesis at long-range. *Oncotarget* 5, 3287–3306. doi: 10.18632/oncotarget.1935

Chiu, L., and Gilbert, S. F. (2015). The birth of the holobiont: multi-species birthing through mutual scaffolding and niche construction. *Biosemiotics* 8, 191–210. doi: 10.1007/s12304-015-9232-5

Chow, R. L., Altmann, C. R., Lang, R. A., and Hemmati-Brivanlou, A. (1999). Pax6 induces ectopic eyes in a vertebrate. *Dev. Suppl.* 126, 4213–4222. doi: 10.1242/dev.126.19.4213

Clark, A., and Chalmers, D. (1998). The extended mind. *Analysis* 58, 7–19.

Cohen-Karni, T., Langer, R., and Kohane, D. S. (2012). The smartest materials: the future of nanoelectronics in medicine. *ACS Nano* 6, 6541–6545. doi: 10.1021/nn302915s

Cook, N. D., Carvalho, G. B., and Damasio, A. (2014). From membrane excitability to metazoan psychology. *Trends Neurosci.* 37, 698–705. doi: 10.1016/j.tins.2014.07.011

Corning, W. C. (1966). Retention of a position discrimination after regeneration in planarians. *Psychanom. Sci.* 5, 17–18.

Corning, W. C. (1967). *Regeneration and Retention of Acquired Information*. Washington, DC: NASA. .

Couzin, I. (2007). Collective minds. *Nature* 445:715. doi: 10.1038/445715a

Couzin, I. D. (2009). Collective cognition in animal groups. *Trends Cogn. Sci.* 13, 36–43. doi: 10.1016/j.tics.2008.10.002

Couzin, I. D., Krause, J., James, R., Ruxton, G. D., and Franks, N. R. (2002). Collective memory and spatial sorting in animal groups. *J. Theor. Biol.* 218, 1–11. doi: 10.1006/jtbi.2002.3065

Damasio, A. R. (2010). *Self Comes to Mind : Constructing the Conscious Brain*, 1st Edn. New York, NY: Pantheon Books, 367.

Damasio, A., and Carvalho, G. B. (2013). The nature of feelings: evolutionary and neurobiological origins. *Nat. Rev. Neurosci.* 14, 143–152. doi: 10.1038/nrn3403

Daniels, B. C., Ellison, C. J., Krakauer, D. C., and Flack, J. C. (2016). Quantifying collectivity. *Curr. Opin. Neurobiol.* 37, 106–113. doi: 10.1016/j.conb.2016.01.012

Danilov, Y., and Tyler, M. (2005). Brainport: an alternative input to the brain. *J. Integr. Neurosci.* 4, 537–550. doi: 10.1142/s0219635205000914

DeMarse, T. B., and Dockendorf, K. P. (2005). "Adaptive flight control with living neuronal networks on microelectrode arrays," in *Proceedings. 2005 IEEE International Joint Conference on Neural Networks*, (Montreal, QC), 1548–1551.

Demarse, T. B., Wagenaar, D. A., Blau, A. W., and Potter, S. M. (2001). The neurally controlled animat: biological brains acting with simulated bodies. *Auton. Robots* 11, 305–310. doi: 10.1023/a:1012407611130

Dennett, D. C. (1987). *The Intentional Stance*. Cambridge, MA: MIT Press, 388.

Dennett, D. C. (1991). *Consciousness Explained*. Boston, MA: Little, Brown and Co.

Dennett, D. C. (2017). *From Bacteria to Bach and Back : The Evolution of Minds*, First Edn. New York, NY: W.W. Norton & Company, 476.

Dexter, J. P., Prabakaran, S., and Gunawardena, J. (2019). A complex hierarchy of avoidance behaviors in a single-cell eukaryote. *Curr. Biol.* 29, 4323–4329.e2. doi: 10.1016/j.cub.2019.10.059

Di Paulo, E. A. (2000). "Homeostatic adaptation to inversion of the visual field and other sensorimotor disruptions," in *Proceedings of the SAB2000 Sixth International Conference on Simulation of Adaptive Behavior : From Animals to Animats*, eds J.-A. Meyer, A. Berthoz, D. Floreano, H. L. Roitblat, and S. W. Wilson, Paris.

di Primio, F., Muller, B. S., and Lengeler, J. W. (2000). "Minimal cognition in unicellular organisms," in *Proceedings of the SAB2000 Sixth International Conference on Simulation of Adaptive Behavior : From Animals to Animats*, eds J.-A. Meyer, A. Berthoz, D. Floreano, H. L. Roitblat, and S. W. Wilson (Paris).

Dietrich, E., Fields, C., Hoffman, D. D., and Prentner, R. (2020). Editorial: epistemic feelings: phenomenology, implementation, and role in cognition. *Front. Psychol.* 11:606046. doi: 10.3389/fpsyg.2020.606046

Ding, S., O'Banion, C. P., Welfare, J. G., and Lawrence, D. S. (2018). Cellular cyborgs: on the precipice of a drug delivery revolution. *Cell Chem. Biol.* 25, 648–658. doi: 10.1016/j.chembiol.2018.03.003

Dobzhansky, T. (1973). Nothing in biology makes sense except in the light of evolution. *Am. Biol. Teach.* 35, 125–129.

Dong, X., Kheiri, S., Lu, Y., Xu, Z., Zhen, M., and Liu, X. (2021). Toward a living soft microrobot through optogenetic locomotion control of Caenorhabditis elegans. *Sci. Robot.* 6:eabe3950. doi: 10.1126/scirobotics.abe3950

Dorahy, M. J., Brand, B. L., Şar, V., Krüger, C., Stavropoulos, P., Martínez-Taboas, A., et al. (2014). Dissociative identity disorder: an empirical overview. *Aust. N. Z. J. Psychiatry* 48, 402–417. doi: 10.1177/0004867414527523

Downing, K. L. (2004). Development and the Baldwin effect. *Artif. Life* 10, 39–63. doi: 10.1162/106454604322875904

Dukas, R. (1998). *Cognitive Ecology: The Evolutionary Ecology of Information Processing and Decision Making*. Chicago, IL: Chicago University Press.

Durant, F., Morokuma, J., Fields, C., Williams, K., Adams, D. S., and Levin, M. (2017). Long-term, stochastic editing of regenerative anatomy via targeting endogenous bioelectric gradients. *Biophys. J.* 112, 2231–2243. doi: 10.1016/j.bpj.2017.04.011

Egeblad, M., Nakasone, E. S., and Werb, Z. (2010). Tumors as organs: complex tissues that interface with the entire organism. *Dev. Cell* 18, 884–901. doi: 10.1016/j.devcel.2010.05.012

Elgart, M., Snir, O., and Soen, Y. (2015). Stress-mediated tuning of developmental robustness and plasticity in flies. *Biochim. Biophys. Acta* 1849, 462–466. doi: 10.1016/j.bbagrm.2014.08.004

Ellis, G. F. R. (2008). On the nature of causation in complex systems. *Transac. R. Soc. South Afr.* 63, 69–84. doi: 10.1080/00359190809519211

Ellis, G. F. R., Noble, D., and O'Connor, T. (2012). Top-down causation: an integrating theme within and across the sciences? Introduction. *Interface Focus* 2, 1–3. doi: 10.1098/Rsfs.2011.0110

Emmons-Bell, M., Durant, F., Hammelman, J., Bessonov, N., Volpert, V., Morokuma, J., et al. (2015). Gap junctional blockade stochastically induces different species-specific head anatomies in genetically wild-type girardia dorotocephala flatworms. *Int. J. Mol. Sci.* 16, 27865–27896. doi: 10.3390/ijms161126065

Emmons-Bell, M., Durant, F., Tung, A., Pietak, A., Miller, K., Kane, A., et al. (2019). Regenerative adaptation to electrochemical perturbation in planaria: a molecular analysis of physiological plasticity. *iScience* 22, 147–165. doi: 10.1016/j.isci.2019.11.014

Epstein, R. (1984). The principle of parsimony and some applications in psychology. *J. Mind. Behav.* 5, 119–130.

Fankhauser, G. (1945a). Maintenance of normal structure in heteroploid salamander larvae, through compensation of changes in cell size by adjustment of cell number and cell shape. *J. Exp. Zool.* 100, 445–455. doi: 10.1002/jez.1401000310

Fankhauser, G. (1945b). The effects of changes in chromosome number on amphibian development. *Q. Rev. Biol.* 20, 20–78. doi: 10.2307/2809003

Feynman, R. (1942). *The Principle of Least Action in Quantum Mechanics*. Ph.D Thesis. Princeton, NJ: Princeton University.

Fields, C., and Levin, M. (2017). Multiscale memory and bioelectric error correction in the cytoplasm–cytoskeleton-membrane system. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 10, e1410. doi: 10.1002/wsbm.1410

Fields, C., Bischof, J., and Levin, M. (2020). Morphological coordination: a common ancestral function unifying neural and non-neural signaling. *Physiology* 35, 16–30. doi: 10.1152/physiol.00027.2019

Fields, C., Hoffman, D. D., Prakash, C., and Singh, M. (2017). Conscious agent networks: formal analysis and application to cognition. 10. *Cogn. Syst. Res.*

Flack, J. C. (2017). Coarse-graining as a downward causation mechanism. *Philos. Trans. A Math. Phys. Eng. Sci.* 375:20160338. doi: 10.1098/rsta.2016.0338

Fontes, P., Komori, J., Lopez, R., Marsh, W., and Lagasse, E. (2020). Development of ectopic livers by hepatocyte transplantation into swine lymph nodes. *Liver Transpl.* 26, 1629–1643. doi: 10.1002/lt.25872

Ford, B. J. (2017). Cellular intelligence: microphenomenology and the realities of being. *Prog. Biophys. Mol. Biol.* 131, 273–287. doi: 10.1016/j.pbiomolbio.2017.08.012

Forraz, N., Wright, K. E., Jurga, M., and McGuckin, C. P. (2013). Experimental therapies for repair of the central nervous system: stem cells and tissue engineering. *J. Tissue Eng. Regen. Med.* 7, 523–536. doi: 10.1002/term.552

Frank, S. A. (2018). Measurement invariance explains the universal law of generalization for psychological perception. *Proc. Natl. Acad. Sci. U.S.A.* 115, 9803–9806. doi: 10.1073/pnas.1809787115

Frank, S. A. (2019a). Evolutionary design of regulatory control. I. A robust control theory analysis of tradeoffs. *J. Theor. Biol.* 463, 121–137. doi: 10.1016/j.jtbi.2018.12.023

Frank, S. A. (2019b). Evolutionary design of regulatory control. II. Robust error-correcting feedback increases genetic and phenotypic variability. *J. Theor. Biol.* 468, 72–81. doi: 10.1016/j.jtbi.2019.02.012

Friston, K. (2013). Life as we know it. *J. R. Soc. Interface* 10:20130475. doi: 10.1098/rsif.2013.0475

Friston, K. J., Shiner, T., FitzGerald, T., Galea, J. M., Adams, R., Brown, H., et al. (2012). Dopamine, affordance and active inference. *PLoS Comput. Biol.* 8:e1002327. doi: 10.1371/journal.pcbi.1002327

Friston, K. J., Stephan, K. E., Montague, R., and Dolan, R. J. (2014). Computational psychiatry: the brain as a phantastic organ. *Lancet Psychiatry* 1, 148–158. doi: 10.1016/S2215-0366(14)70275-5

Friston, K., and Ao, P. (2012). Free energy, value, and attractors. *Comput. Math. Methods Med.* 2012:937860. doi: 10.1155/2012/937860

Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., and Pezzulo, G. (2015b). Active inference and epistemic value. *Cogn. Neurosci.* 6, 187–214. doi: 10.1080/17588928.2015.1020053

Friston, K., Levin, M., Sengupta, B., and Pezzulo, G. (2015a). Knowing one's place: a free-energy approach to pattern regulation. *J. R. Soc. Interface* 12:20141383. doi: 10.1098/rsif.2014.1383

Friston, K., Schwartenbeck, P., Fitzgerald, T., Moutoussis, M., Behrens, T., and Dolan, R. J. (2013). The anatomy of choice: active inference and agency. *Front. Hum. Neurosci.* 7:598. doi: 10.3389/fnhum.2013.00598

Friston, K., Sengupta, B., and Auletta, G. (2014). Cognitive dynamics: from attractors to active inference. *Proc. IEEE* 102, 427–445. doi: 10.1109/Jproc.2014.2306251

Fukumoto, T., Kema, I. P., and Levin, M. (2005b). Serotonin signaling is a very early step in patterning of the left-right axis in chick and frog embryos. *Curr. Biol.* 15, 794–803. doi: 10.1016/j.cub.2005.03.044

Fukumoto, T., Blakely, R., and Levin, M. (2005a). Serotonin transporter function is an early step in left-right patterning in chick and frog embryos. *Dev. Neurosci.* 27, 349–363. doi: 10.1159/000088451

Gao, J., Liu, Y. Y., D'Souza, R. M., and Barabasi, A. L. (2014). Target control of complex networks. *Nat. Commun.* 5:5415. doi: 10.1038/ncomms6415

Gawne, R., McKenna, K. Z., and Levin, M. (2020). Competitive and coordinative interactions between body parts produce adaptive developmental outcomes. *BioEssays* 42:e1900245. doi: 10.1002/bies.201900245

Gazzaniga, M. S. (1970). *The Bisected Brain*. New York, NY: Appleton-Century-Crofts.

Georgiev, G., and Georgiev, I. (2002). The least action and the metric of an organized system. *Open Syst. Inf. Dyn.* 9, 371–380. doi: 10.1023/A:1021858318296

Gerhart, J., and Kirschner, M. (2007). The theory of facilitated variation. *Proc. Natl. Acad. Sci. U.S.A.* 104(Suppl. 1), 8582–8589. doi: 10.1073/pnas.0701035104

Gershman, S. J., Balbi, P. E., Gallistel, C. R., and Gunawardena, J. (2021). Reconsidering the evidence for learning in single cells. *Elife* 10:e61907. doi: 10.7554/eLife.61907.

Ghoneim, M. M., and Block, R. I. (1997). Learning and memory during general anesthesia: an update. *Anesthesiology* 87, 387–410. doi: 10.1097/00000542-199708000-00027

Ginsburg, S., and Jablonka, E. (2021). Evolutionary transitions in learning and cognition. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 376:20190766. doi: 10.1098/rstb.2019.0766

Giselbrecht, S., Rapp, B. E., and Niemeyer, C. M. (2013). The chemistry of cyborgs–interfacing technical devices with organisms. *Angew. Chem. Int. Ed. Engl.* 52, 13942–13957. doi: 10.1002/anie.201307495

Godfrey-Smith, P. (2009). *Darwinian Populations and Natural Selection*. Oxford: Oxford University Press, 207.

Goel, P., and Mehta, A. (2013). Learning theories reveal loss of pancreatic electrical connectivity in diabetes as an adaptive response. *PLoS One* 8:e70366. doi: 10.1371/journal.pone.0070366

Gottlieb, J., Oudeyer, P. Y., Lopes, M., and Baranes, A. (2013). Information-seeking, curiosity, and attention: computational and neural mechanisms. *Trends Cogn. Sci.* 17, 585–593. doi: 10.1016/j.tics.2013.09.001

Green, A. M., and Kalaska, J. F. (2011). Learning to move machines with the mind. *Trends Neurosci.* 34, 61–75. doi: 10.1016/j.tins.2010.11.003

Gremiaux, A., Yokawa, K., Mancuso, S., and Baluska, F. (2014). Plant anesthesia supports similarities between animals and plants: claude Bernard's forgotten studies. *Plant Signal. Behav.* 9:e27886. doi: 10.4161/psb.27886

Grossberg, S. (1978). "Communication, memory, and development," in *Progress in Theoretical Biology*, Vol. 5, eds R. Rosen and F. Snell (New York, NY: Academic Press).

Hadj-Chikh, L. Z., Steele, M. A., and Smallwood, P. D. (1996). Caching decisions by grey squirrels: a test of the handling time and perishability hypotheses. *Anim. Behav.* 52, 941–948. doi: 10.1006/anbe.1996.0242

Haigh, E. L. (1976). Vitalism, the soul, and sensibility: the physiology of Theophile Bordeu. *J. Hist. Med. Allied Sci.* 31, 30–41. doi: 10.1093/jhmas/xxxi.1.30

Hankey, A. (2015). A complexity basis for phenomenology: how information states at criticality offer a new approach to understanding experience of self, being and time. *Prog. Biophys. Mol. Biol.* 119, 288–302. doi: 10.1016/j.pbiomolbio.2015.07.010

Harman, G. (1973). *Thought*. New Jersey, NJ: Princeton.

Harris, M. P. (2021). Bioelectric signaling as a unique regulator of development and regeneration. *Development* 148:dev180794. doi: 10.1242/dev.180794

Heams, T. (2012). Selection within organisms in the nineteenth century: Wilhelm Roux's complex legacy. *Prog. Biophys. Mol. Biol.* 110, 24–33. doi: 10.1016/j.pbiomolbio.2012.04.004

Hernandez-Diaz, S., and Levin, M. (2014). Alteration of bioelectrically-controlled processes in the embryo: a teratogenic mechanism for anticonvulsants. *Reprod. Toxicol.* 47, 111–114. doi: 10.1016/j.reprotox.2014.04.008

Herrera-Delgado, E., Perez-Carrasco, R., Briscoe, J., and Sollich, P. (2018). Memory functions reveal structural properties of gene regulatory networks. *PLoS Comput. Biol.* 14:e1006003. doi: 10.1371/journal.pcbi.1006003

Hoel, E. P., Albantakis, L., and Tononi, G. (2013). Quantifying causal emergence shows that macro can beat micro. *Proc. Natl. Acad. U.S.A.* 110, 19790–19795. doi: 10.1073/pnas.1314922110

Hoel, E. P., Albantakis, L., Marshall, W., and Tononi, G. (2016). Can the macro beat the micro? Integrated information across spatiotemporal scales. *Neurosci. Conscious.* 2016:niw012. doi: 10.1093/nc/niw012

Hoffman, D. D. (2017). "The interface theory of perception," in *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience*, ed. J. T. Wixted (Hoboken, NJ: Wiley).

Hoffman, D. D., Singh, M., and Prakash, C. (2015). The interface theory of perception. *Psychon. Bull. Rev.* 22, 1480–1506. doi: 10.3758/s13423-015-0890-8

Hogenson, G. B. (2001). The Baldwin effect: a neglected influence on C. G. Jung's evolutionary thinking. *J. Anal. Psychol.* 46, 591–611. doi: 10.1111/1465-5922.00269

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. U.S.A.* 79, 2554–2558. doi: 10.1073/pnas.79.8.2554

Hover, S., Foster, B., Barr, J. N., and Mankouri, J. (2017). Viral dependence on cellular ion channels - an emerging anti-viral target? *J. Gen. Virol.* 98, 345–351. doi: 10.1099/jgv.0.000712

Huang, H., Liu, S., and Kornberg, T. B. (2019). Glutamate signaling at cytoneme synapses. *Science* 363, 948–955. doi: 10.1126/science.aat5053

Hughes, D. P., Araujo, J. P., Loreto, R. G., Quevillon, L., de Bekker, C., and Evans, H. C. (2016). From so simple a beginning: the evolution of behavioral manipulation by fungi. *Adv Genet* 94, 437–469. doi: 10.1016/bs.adgen.2016.01.004

Humphries, J., Xiong, L., Liu, J., Prindle, A., Yuan, F., Arjes, H. A., et al. (2017). Species-independent attraction to biofilms through electrical signaling. *Cell* 168, 200–209.e12. doi: 10.1016/j.cell.2016.12.014

Hyun, I., Scharf-Deering, J. C., and Lunshof, J. E. (2020). Ethical issues related to brain organoid research. *Brain Res.* 1732:146653. doi: 10.1016/j.brainres.2020.146653

Inoue, J. (2008). A simple Hopfield-like cellular network model of plant intelligence. *Prog. Brain Res.* 168, 169–174. doi: 10.1016/S0079-6123(07)68014-5

James, W. (1890). *Principles of Psychology*. New York, NY: Henry Holt and Co.

Jekely, G., Keijzer, F., and Godfrey-Smith, P. (2015). An option space for early neural evolution. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370:20150181. doi: 10.1098/rstb.2015.0181

Jennings, H. S. (1906). *Behavior of the Lower Organisms*. New York, NY: The Columbia university press, 366.

Kaila, V. R. I., and Annila, A. (2008). Natural selection for least action. *Proc. R. Soc A* 464, 3055–3070. doi: 10.1098/Rspa.2008.0178

Kang, J. H., Manousaki, T., Franchini, P., Kneitz, S., Schartl, M., and Meyer, A. (2015). Transcriptomics of two evolutionary novelties: how to make a sperm-transfer organ out of an anal fin and a sexually selected "sword" out of a caudal fin. *Ecol. Evol.* 5, 848–864. doi: 10.1002/ece3.1390

Karpas, E. D., Shklarsh, A., and Schneidman, E. (2017). Information socialtaxis and efficient collective behavior emerging in groups of information-seeking agents. *Proc. Natl. Acad. Sci. U.S.A.* 114, 5589–5594. doi: 10.1073/pnas.1618055114

Keijzer, F. (2015). Moving and sensing without input and output: early nervous systems and the origins of the animal sensorimotor organization. *Biol. Philos.* 30, 311–331. doi: 10.1007/s10539-015-9483-1

Keijzer, F., van Duijn, M., and Lyon, P. (2013). What nervous systems do: early evolution, input-output, and the skin brain thesis. *Adapt Behav.* 21, 67–85. doi: 10.1177/1059712312465330

Kelz, M. B., Garcia, P. S., Mashour, G. A., and Solt, K. (2019). Escape from oblivion: neural mechanisms of emergence from general Anesthesia. *Anesth. Analg.* 128, 726–736. doi: 10.1213/ANE.0000000000004006

Koshland, D. E. (1983). The bacterium as a model neuron. *Trends Neurosci.* 6, 133–137. doi: 10.1016/0166-2236(83)90066-8

Kouvaris, K., Clune, J., Kounios, L., Brede, M., and Watson, R. A. (2017). How evolution learns to generalise: using the principles of learning theory to understand the evolution of developmental organisation. *PLoS Comput. Biol.* 13:e1005358. doi: 10.1371/journal.pcbi.1005358

Krakauer, D., Bertschinger, N., Olbrich, E., Ay, N., and Flack, J. C. (2014). *The Information Theory of Individuality. arXiv [Preprint]*. Available online at: https://arxiv.org/abs/1412.2447 (accessed February 2, 2022).

Kralj, J. M., Hochbaum, D. R., Douglass, A. D., and Cohen, A. E. (2011). Electrical spiking in *Escherichia coli* probed with a fluorescent voltage-indicating protein. *Science* 333, 345–348. doi: 10.1126/science.1204763

Kriegman, S., Blackiston, D., Levin, M., and Bongard, J. (2020a). A scalable pipeline for designing reconfigurable organisms. *Proc. Natl. Acad. Sci. U.S.A.* 117, 1853–1859. doi: 10.1073/pnas.1910837117

Kriegman, S., Nasab, A. M., Shah, D., Steele, H., Branin, G., Levin, M., et al. (2020b). "Scalable sim-to-real transfer of soft robot designs," in *Proceedings of the 2020 3rd IEEE International Conference on Soft Robotics (Robosoft)*, (New Haven, CT), 359–366.

Krotov, D. (2021). *Hierarchical Associative Memory*. Available online at: https://ui.adsabs.harvard.edu/abs/2021arXiv210706446K (accessed July 01, 2021).

Kuchling, F., Friston, K., Georgiev, G., and Levin, M. (2020a). Integrating variational approaches to pattern formation into a deeper physics: reply to comments on "Morphogenesis as Bayesian inference: a variational approach to pattern formation and manipulation in complex biological systems". *Phys. Life Rev.* 33, 125–128. doi: 10.1016/j.plrev.2020.07.001

Kuchling, F., Friston, K., Georgiev, G., and Levin, M. (2020b). Morphogenesis as Bayesian inference: a variational approach to pattern formation and control in complex biological systems. *Phys. Life Rev.* 33, 88–108. doi: 10.1016/j.plrev.2019.06.001

Lan, G., and Tu, Y. (2016). Information processing in bacteria: memory, computation, and statistical physics: a key issues review. *Rep. Prog. Phys.* 79:052601. doi: 10.1088/0034-4885/79/5/052601

Langton, C. G. (1995). *Artificial Life: An Overview*. Cambridge, MA: MIT Press.

Larkin, J. W., Zhai, X., Kikuchi, K., Redford, S. E., Prindle, A., Liu, J., et al. (2018). Signal percolation within a bacterial community. *Cell Syst* 7, 137–145.e3. doi: 10.1016/j.cels.2018.06.005

Law, R., and Levin, M. (2015). Bioelectric memory: modeling resting potential bistability in amphibian embryos and mammalian cells. *Theor. Biol. Med. Model* 12:22. doi: 10.1186/s12976-015-0019-9

Leithe, E., Sirnes, S., Omori, Y., and Rivedal, E. (2006). Downregulation of gap junctions in cancer cells. *Crit. Rev. Oncog.* 12, 225–256. doi: 10.1615/critrevoncog.v12.i3-4.30

Levin, M. (2011). The wisdom of the body: future techniques and approaches to morphogenetic fields in regenerative medicine, developmental biology and cancer. *Regen. Med.* 6, 667–673. doi: 10.2217/rme.11.69

Levin, M. (2019). The computational boundary of a "Self": developmental bioelectricity drives multicellularity and scale-free cognition. *Front. Psychol.* 10:2688. doi: 10.3389/fpsyg.2019.02688

Levin, M. (2020). Life, death, and self: fundamental questions of primitive cognition viewed through the lens of body plasticity and synthetic organisms. *Biochem. Biophys. Res. Commun.* 564, 114–133. doi: 10.1016/j.bbrc.2020.10.077

Levin, M. (2021a). Bioelectric signaling: reprogrammable circuits underlying embryogenesis, regeneration, and cancer. *Cell* 184, 1971–1989. doi: 10.1016/j.cell.2021.02.034

Levin, M. (2021b). Bioelectrical approaches to cancer as a problem of the scaling of the cellular self. *Prog. Biophys. Mol. Biol.* 165, 102–113. doi: 10.1016/j.pbiomolbio.2021.04.007

Levin, M. (2022). TAME: technological approach to mind everywhere. *PsyArXiv [Preprint]* doi: 10.31234/osf.io/t6e8p

Levin, M., and Dennett, D. C. (2020). *Cognition All the Way Down*. Melbourne: Aeon.

Levin, M., and Martyniuk, C. J. (2018). The bioelectric code: an ancient computational medium for dynamic control of growth and form. *Biosystems* 164, 76–93. doi: 10.1016/j.biosystems.2017.08.009

Levin, M., Bongard, J., and Lunshof, J. E. (2020). Applications and ethics of computer-designed organisms. *Nat. Rev. Mol. Cell Biol.* 21, 655–656. doi: 10.1038/s41580-020-00284-z

Levin, M., Buznikov, G. A., and Lauder, J. M. (2006). Of minds and embryos: left-right asymmetry and the serotonergic controls of pre-neural morphogenesis. *Dev. Neurosci.* 28, 171–185. doi: 10.1159/000091915

Levin, M., Keijzer, F., Lyon, P., and Arendt, D. (2021). Uncovering cognitive similarities and differences, conservation and innovation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 376:20200458. doi: 10.1098/rstb.2020.0458

Levin, M., Pezzulo, G., and Finkelstein, J. M. (2017). Endogenous bioelectric signaling networks: exploiting voltage gradients for control of growth and form. *Annu. Rev. Biomed. Eng.* 19, 353–387. doi: 10.1146/annurev-bioeng-071114-040647

Li, W. L., Matsuhisa, N., Liu, Z. Y., Wang, M., Luo, Y. F., Cai, P. Q., et al. (2021). An on-demand plant-based actuator created using conformable electrodes. *Nat. Electron.* 4, 134–142. doi: 10.1038/s41928-020-00530-4

Liebeskind, B. J., Hillis, D. M., and Zakon, H. H. (2015). Convergence of ion channel genome content in early animal evolution. *Proc. Natl. Acad. Sci. U.S.A.* 112, E846–E851. doi: 10.1073/pnas.1501195112

Liu, J., Martinez-Corral, R., Prindle, A., Lee, D. D., Larkin, J., Gabalda-Sagarra, M., et al. (2017). Coupling between distant biofilms and emergence of nutrient time-sharing. *Science* 356, 638–642. doi: 10.1126/science.aah4204

Liu, Y. Y., Slotine, J. J., and Barabasi, A. L. (2011). Controllability of complex networks. *Nature* 473, 167–173. doi: 10.1038/nature10011

Lobo, D., Solano, M., Bubenik, G. A., and Levin, M. (2014). A linear-encoding model explains the variability of the target morphology in regeneration. *J. R. Soc.* 11:20130918. doi: 10.1098/rsif.2013.0918

Lowell, J., and Pollack, J. (eds) (2016). "Developmental encodings promote the emergence of hierarchical modularity," in *Proceedings of the Artificial Life Conference 2016*, (Cancun: MIT Press).

Lyon, P. (2006). The biogenic approach to cognition. *Cogn. Process.* 7, 11–29. doi: 10.1007/s10339-005-0016-8

Lyon, P., and Kuchling, F. (2021). Valuing what happens: a biogenic approach to valence and (potentially) affect. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 376:20190752. doi: 10.1098/rstb.2019.0752

Lyon, P., Keijzer, F., Arendt, D., and Levin, M. (2021). Reframing cognition: getting down to biological basics. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 376:20190750. doi: 10.1098/rstb.2019.0750

Man, K., and Damasio, A. (2019). Homeostasis and soft robotics in the design of feeling machines. *Nat. Mach. Intell.* 1, 446–452. doi: 10.1038/s42256-019-0103-7

Manicka, S., and Harvey, I. (eds) (2008). *'Psychoanalysis' of a Minimal Agent. Artificial Life XI*; Winchester, UK.

Manicka, S., and Levin, M. (2019b). The cognitive lens: a primer on conceptual tools for analysing information processing in developmental and regenerative morphogenesis. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 374:20180369. doi: 10.1098/rstb.2018.0369

Manicka, S., and Levin, M. (2019a). Modeling somatic computation with non-neural bioelectric networks. *Sci. Re.p* 9:18612. doi: 10.1038/s41598-019-54859-8

Mar, R. A., Kelley, W. M., Heatherton, T. F., and Macrae, C. N. (2007). Detecting agency from the biological motion of veridical vs animated agents. *Soc. Cogn. Affect. Neurosci.* 2, 199–205. doi: 10.1093/scan/nsm011

Marr, D. (1982). *Vision : A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco, CA: W.H. Freeman, 397.

Martinez-Corral, R., Liu, J., Prindle, A., Suel, G. M., and Garcia-Ojalvo, J. (2019). Metabolic basis of brain-like electrical signalling in bacterial communities. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 374:20180382. doi: 10.1098/rstb.2018.0382

Martinez-Corral, R., Liu, J., Suel, G. M., and Garcia-Ojalvo, J. (2018). Bistable emergence of oscillations in growing *Bacillus subtilis* biofilms. *Proc. Natl. Acad. Sci. U.S.A.* 115, E8333–E8340. doi: 10.1073/pnas.1805004115

Maslow, A. H. (1943). A theory of human motivation. *Psychol. Rev.* 50, 370–396.

Maturana, H. R., and Varela, F. J. (1980). *Autopoiesis and Cognition : The Realization of the Living*. Dordrecht: D. Reidel Pub. Co, 141.

Maynard Smith, J. (1999). *Shaping Life : Genes, Embryos, and Evolution*. New Haven, CT: Yale University Press, 50.

Maynard Smith, J., and Szathmáry, E. (1995). *The Major Transitions in Evolution*. New York, NY: W.H. Freeman Spektrum, 346.

Mayr, E. (1992). The idea of teleology. *J. Hist. Ideas* 53, 117–135. doi: 10.2307/2709913

McConnell, J. V., and Shelby, J. M. (1970). "Memory transfer experiments in invertebrates," in *Molecular Mechanisms in Memory and Learning*, ed. G. Ungar (New York, NY: Plenum Press), 71–101.

McConnell, J. V., Jacobson, A. L., and Kimble, D. P. (1959). The effects of regeneration upon retention of a conditioned response in the planarian. *J. Comp. Physiol. Psychol.* 52, 1–5. doi: 10.1037/h0048028

McEvoy, J. W. (2009). Evolutionary game theory: lessons and limitations, a cancer perspective. *Br. J. Cancer* 101, 2060–1;author reply 2062–3. doi: 10.1038/sj.bjc.6605444

McEwen, B. S. (1998). Stress, adaptation, and disease. allostasis and allostatic load. *Ann. N.Y. Acad. Sci.* 840, 33–44. doi: 10.1111/j.1749-6632.1998.tb09546.x

McLaughlin, K. A., and Levin, M. (2018). Bioelectric signaling in regeneration: mechanisms of ionic controls of growth and form. *Dev. Biol.* 433, 177–189. doi: 10.1016/j.ydbio.2017.08.032

McNamara, H. M., Salegame, R., Tanoury, Z. A., Xu, H., Begum, S., Ortiz, G., et al. (2020). Bioelectrical domain walls in homogeneous tissues. *Nat. Phys.* 16, 357–364. doi: 10.1038/s41567-019-0765-4

McNamara, H. M., Salegame, R., Tanoury, Z. A., Xu, H., Begum, S., Ortiz, G., et al. (2019). Bioelectrical signaling via domain wall migration. *bioRxiv [Preprint]* 570440. doi: 10.1101/570440

McNamara, H. M., Zhang, H., Werley, C. A., and Cohen, A. E. (2016). Optically controlled oscillators in an engineered bioelectric tissue. *Phys. Rev. X* 6:031001.

McShea, D. W. (2012). Upper-directed systems: a new approach to teleology in biology. *Biol. Philos.* 27, 663–684. doi: 10.1007/s10539-012-9326-2

McShea, D. W. (2013). Machine wanting. *Stud. Hist. Philos. Biol. Biomed. Sci.* 44(4 Pt. B), 679–687. doi: 10.1016/j.shpsc.2013.05.015

McShea, D. W. (2016). Freedom and purpose in biology. *Stud. Hist. Philos. Biol. Biomed. Sci.* 58, 64–72. doi: 10.1016/j.shpsc.2015.12.002

Mehrali, M., Bagherifard, S., Akbari, M., Thakur, A., Mirani, B., Mehrali, M., et al. (2018). Blending electronics with the human body: a pathway toward a cybernetic future. *Adv. Sci.* 5:1700931. doi: 10.1002/advs.201700931

Melo, D., Porto, A., Cheverud, J. M., and Marroig, G. (2016). Modularity: genes, development and evolution. *Annu. Rev. Ecol. Evol. Syst.* 47, 463–486. doi: 10.1146/annurev-ecolsys-121415-032409

Merritt, T., Hamidi, F., Alistar, M., and DeMenezes, M. (2020). Living media interfaces: a multi-perspective analysis of biological materials for interaction. *Digit Creat.* 31, 1–21. doi: 10.1080/14626268.2019.1707231

Michod, R. E., and Nedelcu, A. M. (2003). On the reorganization of fitness during evolutionary transitions in individuality. *Integr. Comp. Biol.* 43, 64–73. doi: 10.1093/icb/43.1.64

Mikhalevich, I., and Powell, R. (eds) (2020). Minds without spines: evolutionarily inclusive animal ethics. *Anim. Sentience* 29:2020.

Miller, S. D., and Triggiano, P. J. (1992). The psychophysiological investigation of multiple personality disorder: review and update. *Am. J. Clin. Hypn.* 35, 47–61. doi: 10.1080/00029157.1992.10402982

Montgomery, B. A. (2003). *Consciousness and Personhood in Split-Brain Patients: Dissertation*. Oklahoma: The University of Oklahoma, 1–231.

Moran, Y., Barzilai, M. G., Liebeskind, B. J., and Zakon, H. H. (2015). Evolution of voltage-gated ion channels at the emergence of Metazoa. *J. Exp. Biol.* 218(Pt. 4), 515–525. doi: 10.1242/jeb.110270

Morgan, C. L. (1903). "Other minds than ours," in *An Introduction to Comparative Psychology*, ed. W. Scott (London: Walter Scott Publishing), 36–59.

Morgan, T. H. (1904). The control of heteromorphosis in Planaria maculata. *Arch. Für Entw. Mech.* 17, 683–694.

Muller, F. J., and Schuppert, A. (2011). Few inputs can reprogram biological networks. *Nature* 478, E4;discussion E4–5. doi: 10.1038/nature10543

Nagel, E. (1979). *Teleology Revisited and Other Essays in the Philosophy and History of Science*. New York, NY: Columbia University Press, 352.

Nagel, T. (1971). Brain bisection and the unity of consciousness. *Synthese* 22, 396–413.

Nagel, T. (1974). What is it like to be a bat? *Philos. Rev.* 83, 435–450. doi: 10.1111/1468-5930.00141

Nasuto, S. J., and Hayashi, Y. (2016). Anticipation: beyond synthetic biology and cognitive robotics. *Biosystems* 148, 22–31. doi: 10.1016/j.biosystems.2016.07.011

Nicolis, S. C., Zabzina, N., Latty, T., and Sumpter, D. J. (2011). Collective irrationality and positive feedback. *PLoS One* 6:e18901. doi: 10.1371/journal.pone.0018901

Noble, D. (2010). Biophysics and systems biology. *Philos. Trans. A Math. Phys. Eng. Sci.* 368, 1125–1139. doi: 10.1098/rsta.2009.0245

Noble, D. (2011). The aims of systems biology: between molecules and organisms. *Pharmacopsychiatry* 44(Suppl. 1), S9–S14. doi: 10.1055/s-0031-1271703

Noble, D. (2012). A theory of biological relativity: no privileged level of causation. *Interface Focus* 2, 55–64. doi: 10.1098/Rsfs.2011.0067

Nogi, T., and Levin, M. (2005). Characterization of innexin gene expression and functional roles of gap-junctional communication in planarian regeneration. *Dev. Biol.* 287, 314–335. doi: 10.1016/j.ydbio.2005.09.002

Norman, T. M., Lord, N. D., Paulsson, J., and Losick, R. (2013). Memory and modularity in cell-fate decision making. *Nature* 503, 481–486. doi: 10.1038/nature12804

Ogborn, J., Hanc, J., and Taylor, E. (eds) (2006). "Action on stage: historical introduction," in *Proceedings of the GIREP Conference, Modeling in Physics and Physics Education*, (Amsterdam: AMSTEL Institute).

Orive, G., Taebnia, N., and Dolatshahi-Pirouz, A. A. (2020). New era for cyborg science is emerging: the promise of cyborganic beings. *Adv. Healthc. Mater.* 9:e1901023. doi: 10.1002/adhm.201901023

Otopalik, A. G., Sutton, A. C., Banghart, M., and Marder, E. (2017). When complex neuronal structures may not matter. *Elife* 6:e23508. doi: 10.7554/eLife.23508

Oudeyer, P. Y., and Kaplan, F. (2007). What is intrinsic motivation? A typology of computational approaches. *Front. Neurorobot.* 1:6. doi: 10.3389/neuro.12.006.2007

Oudeyer, P.-Y., and Kaplan, F. (2013). *How Can We Define Intrinsic Motivation*. Available online at: http://www.pyoudeyer.com/epirob08OudeyerKaplan.pdf

Oviedo, N. J., Morokuma, J., Walentek, P., Kema, I. P., Gu, M. B., Ahn, J. M., et al. (2010). Long-range neural and gap junction protein-mediated cues control polarity during planarian regeneration. *Dev. Biol.* 339, 188–199. doi: 10.1016/j.ydbio.2009.12.012

Pacheco, J. M., Santos, F. C., and Dingli, D. (2014). The ecology of cancer from an evolutionary game theory perspective. *Interface Focus* 4:20140019. doi: 10.1098/rsfs.2014.0019

Pai, V. P., Aw, S., Shomrat, T., Lemire, J. M., and Levin, M. (2012). Transmembrane voltage potential controls embryonic eye patterning in Xenopus laevis. *Development* 139, 313–323. doi: 10.1242/dev.073759

Pai, V. P., Cervera, J., Mafe, S., Willocq, V., Lederer, E. K., and Levin, M. (2020). HCN2 channel-induced rescue of brain teratogenesis via local and long-range bioelectric repair. *Front. Cell Neurosci.* 14:136. doi: 10.3389/fncel.2020.00136

Pai, V. P., Pietak, A., Willocq, V., Ye, B., Shi, N. Q., and Levin, M. (2018). HCN2 rescues brain defects by enforcing endogenous voltage pre-patterns. *Nat. Commun.* 9:998. doi: 10.1038/s41467-018-03334-5

Pai, V. P., Willocq, V., Pitcairn, E. J., Lemire, J. M., Pare, J. F., Shi, N. Q., et al. (2017). HCN4 ion channel function is required for early events that regulate anatomical left-right patterning in a nodal and lefty asymmetric gene expression-independent manner. *Biol. Open* 6, 1445–1457. doi: 10.1242/bio.025957

Peters, A., McEwen, B. S., and Friston, K. (2017). Uncertainty and stress: why it causes diseases and how it is mastered by the brain. *Prog. Neurobiol.* 156, 164–188. doi: 10.1016/j.pneurobio.2017.05.004

Pezzulo, G., and 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.* 7, 1487–1517. doi: 10.1039/c5ib00221d

Pezzulo, G., and Levin, M. (2016). Top-down models in biology: explanation and control of complex living systems above the molecular level. *J. R. Soc. Interface* 13:20160555. doi: 10.1098/rsif.2016.0555

Pezzulo, G., Lapalme, J., Durant, F., and Levin, M. (2021). Bistability of somatic pattern memories: stochastic outcomes in bioelectric circuits underlying regeneration. *Philos. Proc. R. Soc. B* 376:20190765. doi: 10.1098/rstb.2019.0765

Pietak, A., and Levin, M. (2017). Bioelectric gene and reaction networks: computational modelling of genetic, biochemical and bioelectrical dynamics in pattern regulation. *J. R. Soc. Interface* 14:20170425. doi: 10.1098/rsif.2017.0425

Pietak, A., and Levin, M. (2018). Bioelectrical control of positional information in development and regeneration: a review of conceptual and computational advances. *Prog. Biophys. Mol. Biol.* 137, 52–68. doi: 10.1016/j.pbiomolbio.2018.03.008

Pietsch, P., and Schneider, C. W. (1969). Brain transplantation in salamanders - an approach to memory transfer. *Brain Res.* 14, 707–715. doi: 10.1016/0006-8993(69)90210-8

Pinet, K., Deolankar, M., Leung, B., and McLaughlin, K. A. (2019). Adaptive correction of craniofacial defects in pre-metamorphic Xenopus laevis tadpoles involves thyroid hormone-independent tissue remodeling. *Development* 146:dev175893. doi: 10.1242/dev.175893

Pio-Lopez, L. (2021). The rise of the biocyborg: synthetic biology, artificial chimerism and human enhancement. *N. Genet. Soc.* 40, 599–619. doi: 10.1080/14636778.2021.2007064

Pitcairn, E., Harris, H., Epiney, J., Pai, V. P., Lemire, J. M., Ye, B., et al. (2017). Coordinating heart morphogenesis: a novel role for Hyperpolarization-activated cyclic nucleotide-gated (HCN) channels during cardiogenesis in Xenopus laevis. *Commun. Integr. Biol.* 10:e1309488. doi: 10.1080/19420889.2017.1309488

Pittendrigh, C. S. (1958). "Adaptation, natural selection, and behavior," in *Behavior and Evolution*, eds A. Roe and G. G. Simpson (New Haven, CT: Yale University Press), 390–416.

Posfai, M., Gao, J., Cornelius, S. P., Barabasi, A. L., and D'Souza, R. M. (2016). Controllability of multiplex, multi-time-scale networks. *Phys. Rev. E* 94:032316. doi: 10.1103/PhysRevE.94.032316

Potter, S. M., Wagenaar, D. A., and DeMarse, T. B. (eds) (2005). "Closing the loop: stimulation feedback systems for embodied MEA cultures," in *Advances in Network Electrophysiology Using Multi-Electrode Arrays*, eds M. Taketani and M. Baudry (New York, NY: Springer). doi: 10.3389/neuro.12.005.2007

Potter, S. M., Wagenaar, D. A., Madhavan, R., and DeMarse, T. B. (2003). "Long-term bidirectional neuron interfaces for robotic control, and in vitro learning studies," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Vol. 25, (Piscataway, NJ: IEEE), 3690–3693. doi: 10.1109/Iembs.2003.1280959

Power, D. A., Watson, R. A., Szathmary, E., Mills, R., Powers, S. T., Doncaster, C. P., et al. (2015). What can ecosystems learn? Expanding evolutionary ecology with learning theory. *Biol. Direct.* 10:69. doi: 10.1186/s13062-015-0094-1

Prakash, C., Fields, C., Hoffman, D. D., Prentner, R., and Singh, M. (2020). Fact, fiction, and fitness. *Entropy* 22:514. doi: 10.3390/e22050514

Prentner, R. (2019). Consciousness and topologically structured phenomenal spaces. *Conscious. Cogn.* 70, 25–38. doi: 10.1016/j.concog.2019.02.002

Prindle, A., Liu, J., Asally, M., Ly, S., Garcia-Ojalvo, J., and Suel, G. M. (2015). Ion channels enable electrical communication in bacterial communities. *Nature* 527, 59–63. doi: 10.1038/nature15709

Prinz, A. A., Bucher, D., and Marder, E. (2004). Similar network activity from disparate circuit parameters. *Nat. Neurosci.* 7, 1345–1352. doi: 10.1038/nn1352

Ptito, M., Moesgaard, S. M., Gjedde, A., and Kupers, R. (2005). Cross-modal plasticity revealed by electrotactile stimulation of the tongue in the congenitally blind. *Brain* 128(Pt. 3), 606–614. doi: 10.1093/brain/awh380

Qadri, M. A., and Cook, R. G. (2017). Pigeons and humans use action and pose information to categorize complex human behaviors. *Vision Res.* 131, 16–25. doi: 10.1016/j.visres.2016.09.011

Queller, D. C., and Strassmann, J. E. (2009). Beyond society: the evolution of organismality. *Philos Trans R Soc Lond B Biol Sci* 364, 3143–3155. doi: 10.1098/rstb.2009.0095

Rabinovich, M. I., Huerta, R., Varona, P., and Afraimovich, V. S. (2008). Transient cognitive dynamics, metastability, and decision making. *PLoS Comput. Biol.* 4:e1000072. doi: 10.1371/journal.pcbi.1000072

Raby, C. R., and Clayton, N. S. (2009). Prospective cognition in animals. *Behav. Process.* 80, 314–324. doi: 10.1016/j.beproc.2008.12.005

Raible, D. W., and Ragland, J. W. (2005). Reiterated Wnt and BMP signals in neural crest development. *Semin. Cell Dev. Biol.* 16, 673–682. doi: 10.1016/j.semcdb.2005.06.008

Ramstead, M. J. D., Constant, A., Badcock, P. B., and Friston, K. J. (2019). Variational ecology and the physics of sentient systems. *Phys. Life Rev.* 31, 188–205. doi: 10.1016/j.plrev.2018.12.002

Ray, S. (1999). Survival of olfactory memory through metamorphosis in the fly Musca domestica. *Neurosci. Lett.* 259, 37–40. doi: 10.1016/s0304-3940(98)00892-1

Reber, A. S., and Baluska, F. (2021). Cognition in some surprising places. *Biochem. Biophys. Res. Commun.* 564, 150–157. doi: 10.1016/j.bbrc.2020.08.115

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. doi: 10.1162/106454600300103656

Reid, C. R., Beekman, M., Latty, T., and Dussutour, A. (2013). Amoeboid organism uses extracellular secretions to make smart foraging decisions. *Behav. Ecol.* 24, 812–818. doi: 10.1093/beheco/art032

Reid, C. R., Latty, T., Dussutour, A., and Beekman, M. (2012). Slime mold uses an externalized spatial "memory" to navigate in complex environments. *Proce. Natl. Acad. Sci. U.S.A.* 109, 17490–17494. doi: 10.1073/pnas.1215037109

Reinders, A. A. T. S., Chalavi, S., Schlumpf, Y. R., Vissia, E. M., Nijenhuis, E. R. S., Jäncke, L., et al. (2018). Neurodevelopmental origins of abnormal cortical morphology in dissociative identity disorder. *Acta Psychiatr. Scand.* 137, 157–170. doi: 10.1111/acps.12839

Reinders, A. A. T. S., Marquand, A. F., Schlumpf, Y. R., Chalavi, S., Vissia, E. M., Nijenhuis, E. R. S., et al. (2019). Aiding the diagnosis of dissociative identity disorder: pattern recognition study of brain biomarkers. *Br. J. Psychiatry* 215, 536–544. doi: 10.1192/bjp.2018.255

Ricotti, L., Trimmer, B., Feinberg, A. W., Raman, R., Parker, K. K., Bashir, R., et al. (2017). Biohybrid actuators for robotics: A review of devices actuated by living cells. *Sci Robot.* 2:eaaq0495. doi: 10.1126/scirobotics.aaq0495

Robinson, G. E., and Barron, A. B. (2017). Epigenetics and the evolution of instincts. *Science* 356, 26–27. doi: 10.1126/science.aam6142

Robinson, K. R., and Messerli, M. A. (1996). "Electric embryos: the embryonic epithelium as a generator of developmental information," in *Nerve Growth and Guidance*, ed. C. D. McCaig (London: Portland Press), 131–150.

Rolston, J. D., Gross, R. E., and Potter, S. M. (2009a). A low-cost multielectrode system for data acquisition enabling real-time closed-loop processing with rapid recovery from stimulation artifacts. *Front. Neuroeng.* 2:12. doi: 10.3389/neuro.16.012.2009

Rolston, J. D., Gross, R. E., and Potter, S. M. (2009b). "NeuroRighter: closed-loop multielectrode stimulation and recording for freely moving animals and cell cultures," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Conference*, Vol. 2009, (Minneapolis, MN), 6489–6492. doi: 10.1109/IEMBS.2009.5333589.

Rosen, R. (1973). Dynamical realization of (M,R)-Systems. *Bull. Math. Biol.* 35, 1–9. doi: 10.1007/BF02558788

Rosen, R. (1979). Anticipatory systems in retrospect and prospect. *Gen. Syst.* 24, 11–23.

Rosen, R. (1985). *Anticipatory Systems : Philosophical, Mathematical, and Methodological Foundations*, 1st Edn. New York, NY: Pergamon Press, 436.

Rosenblueth, A., Wiener, N., and Bigelow, J. (1943). Behavior, purpose, and teleology. *Philos. Sci.* 10, 18–24.

Rosser, A., and Svendsen, C. N. (2014). Stem cells for cell replacement therapy: a therapeutic strategy for HD? *Mov. Disord.* 29, 1446–1454. doi: 10.1002/mds.26026

Ruud, G. (1929). Heteronom-orthotopische transplantationen von extremitätenanlagen bei axolotlembryonen. *Wilhelm. Roux Arch. Entwickl. Mech. Org.* 118, 308–351.

Sadoc, J. F., and Mosseri, R. (2007). *Geometrical Frustration*. Cambridge, MA: Cambridge University Press.

Saha, D., Mehta, D., Altan, E., Chandak, R., Traner, M., Lo, R., et al. (2020). Explosive sensing with insect-based biorobots. *Biosens. Bioelectronics: X* 6:100050. doi: 10.1016/j.biosx.2020.100050

Saniova, B., Drobny, M., and Sulaj, M. (2009). Delirium and postoperative cognitive dysfunction after general anesthesia. *Med. Sci. Monit.* 15, CS81–CS87.

Śāntideva Bstan 'dzin rgya m, and Comité de traduction Padmakara (2006). *The Way of the Bodhisattva : a Translation of the Bodhicharyāvatāra*, 2nd Edn. Boston, MA: Shambhala, 222. Distributed in the United States by Random House.

Sasaki, T., and Biro, D. (2017). Cumulative culture can emerge from collective intelligence in animal groups. *Nat. Commun.* 8:15049. doi: 10.1038/ncomms15049

Schlosser, G. (1998). Self-re-production and functionality - A systems-theoretical approach to teleological explanation. *Synthese* 116, 303–354. doi: 10.1023/A:1005073307193

Schlosser, G., and Wagner, G. P. (2004). *Modularity in Development and Evolution*. Chicago, IL: University of Chicago Press, 600.

Schreier, H. I., Soen, Y., and Brenner, N. (2017). Exploratory adaptation in large random networks. *Nat. Commun.* 8:14826. doi: 10.1038/ncomms14826

Schulkin, J., and Sterling, P. (2019). Allostasis: a brain-centered, predictive mode of physiological regulation. *Trends Neurosci.* 42, 740–752. doi: 10.1016/j.tins.2019.07.010

Schwitzgebel, E. (2015). If materialism is true, the United States is probably conscious. *Philos. Stud.* 172, 1697–1721. doi: 10.1007/s11098-014-0387-8

Serre, N. B. C., Kralik, D., Yun, P., Slouka, Z., Shabala, S., and Fendrych, M. (2021). AFB1 controls rapid auxin signalling through membrane depolarization in *Arabidopsis thaliana* root. *Nat. Plants* 7, 1229–1238. doi: 10.1038/s41477-021-00969-z

Sheiman, I. M., and Tiras, K. L. (1996). "Memory and morphogenesis in planaria and beetle," in *Russian Contributions to Invertebrate Behavior*, eds C. I. Abramson, Z. P. Shuranova, and Y. M. Burmistrov (Westport, CT: Praeger), 43–76.

Shimbo, K., Brassard, D. L., Lamb, R. A., and Pinto, L. H. (1996). Ion selectivity and activation of the M2 ion channel of influenza virus. *Biophys. J.* 70, 1335–1346. doi: 10.1016/S0006-3495(96)79690-X

Shoemaker, S. S. (1959). Personal identity and memory. *J. Philosophy* 56, 868–882. doi: 10.2307/2022317

Shomrat, T., and Levin, M. (2013). An automated training paradigm reveals long-term memory in planarians and its persistence through head regeneration. *J. Exp. Biol.* 216(Pt. 20), 3799–3810. doi: 10.1242/jeb.087809

Sims, M. (2020). How to count biological minds: symbiosis, the free energy principle, and reciprocal multiscale integration. *Synthese* 199, 2157–2179. doi: 10.1007/s11229-020-02876-w

Smith, B. P., and Litchfield, C. A. (2010). How well do dingoes, *Canis dingo*, perform on the detour task? *Anim. Behav.* 80, 155–162. doi: 10.1016/j.anbehav.2010.04.017

Soen, Y., Knafo, M., and Elgart, M. (2015). A principle of organization which facilitates broad Lamarckian-like adaptations by improvisation. *Biol. Direct.* 10:68. doi: 10.1186/s13062-015-0097-y

Sole, R., Moses, M., and Forrest, S. (2019). Liquid brains, solid brains. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 374:20190040. doi: 10.1098/rstb.2019.0040

Sordillo, A., and Bargmann, C. I. (2021). Behavioral control by depolarized and hyperpolarized states of an integrating neuron. *Elife* 10:e67723. doi: 10.7554/eLife.67723

Spemann, H. (1967). *Embryonic Development and Induction*. New Haven, CT: Yale University Press.

Spencer, G. J., and Genever, P. G. (2003). Long-term potentiation in bone–a role for glutamate in strain-induced cellular memory? *BMC Cell Biol.* 4:9. doi: 10.1186/1471-2121-4-9

Srivastava, P., Kane, A., Harrison, C., and Levin, M. A. (2020). Meta-analysis of bioelectric data in cancer, embryogenesis, and regeneration. *Bioelectricity* 3, 42–67. doi: 10.1089/bioe.2019.0034

Stockwell, S. R., Landry, C. R., and Rifkin, S. A. (2015). The yeast galactose network as a quantitative model for cellular memory. *Mol. Biosyst.* 11, 28–37. doi: 10.1039/c4mb00448e

Stratford, J. P., Edwards, C. L. A., Ghanshyam, M. J., Malyshev, D., Delise, M. A., Hayashi, Y., et al. (2019). Electrically induced bacterial membrane-potential dynamics correspond to cellular proliferation capacity. *Proc. Natl. Acad. Sci. U.S.A.* 116, 9552–9557. doi: 10.1073/pnas.1901788116

Sullivan, K. G., Emmons-Bell, M., and Levin, M. (2016). Physiological inputs regulate species-specific anatomy during embryogenesis and regeneration. *Commun. Integr. Biol.* 9:e1192733. doi: 10.1080/19420889.2016.1192733

Szilagyi, A., Szabo, P., Santos, M., and Szathmary, E. (2020). Phenotypes to remember: evolutionary developmental memory capacity and robustness. *PLoS Comput. Biol.* 16:e1008425. doi: 10.1371/journal.pcbi.1008425

Tamori, Y., and Deng, W. M. (2014). Compensatory cellular hypertrophy: the other strategy for tissue homeostasis. *Trends Cell Biol.* 24, 230–237. doi: 10.1016/j.tcb.2013.10.005

Tanna, T., and Sachan, V. (2014). Mesenchymal stem cells: potential in treatment of neurodegenerative diseases. *Curr. Stem Cell Res. Ther.* 9, 513–521. doi: 10.2174/1574888x09666140923101110

Taormina, R. J., and Gao, J. H. (2013). Maslow and the motivation hierarchy: measuring satisfaction of the needs. *Am. J. Psychol.* 126, 155–177. doi: 10.5406/amerjpsyc.126.2.0155

Thierry, B., Theraulaz, G., Gautier, J. Y., and Stiegler, B. (1995). Joint memory. *Behav. Process.* 35, 127–140. doi: 10.1016/0376-6357(95)00039-9

Thornton, C. (2017). Predictive processing simplified: the infotropic machine. *Brain Cogn.* 112, 13–24. doi: 10.1016/j.bandc.2016.03.004

Timsit, Y., and Gregoire, S. P. (2021). Towards the idea of molecular brains. *Int. J. Mol. Sci.* 22:11868. doi: 10.3390/ijms222111868

Trewavas, A. J., and Baluska, F. (2011). The ubiquity of consciousness. *EMBO Rep.* 12, 1221–1225. doi: 10.1038/embor.2011.218

Tseng, A. S., Beane, W. S., Lemire, J. M., Masi, A., and Levin, M. (2010). Induction of vertebrate regeneration by a transient sodium current. *J. Neurosci.* 30, 13192–13200. doi: 10.1523/JNEUROSCI.3315-10.2010

Tseng, A., and Levin, M. (2013). Cracking the bioelectric code: Probing endogenous ionic controls of pattern formation. *Commun. Integr. Biol.* 6:e22595. doi: 10.4161/cib.22595

Tsuda, S., Artmann, S., and Zauner, K.-P. (2009). "The phi-bot: a robot controlled by a slime mould," in *Artificial Life Models in Hardware*, eds A. Adamatzky and M. Komosinski (London: Springer), 213–232.

Tully, T., Cambiazo, V., and Kruse, L. (1994). Memory through metamorphosis in normal and mutant Drosophila. *J. Neurosci.* 14, 68–74. doi: 10.1523/JNEUROSCI.14-01-00068.1994

Turner, C. H., Robling, A. G., Duncan, R. L., and Burr, D. B. (2002). Do bone cells behave like a neuronal network? *Calcif. Tissue Int.* 70, 435–442. doi: 10.1007/s00223-001-1024-z

Turner, J. S. (2000). *The Extended Organism : The Physiology of Animal-Built Structures*. Cambridge, MA: Harvard University Press, 235.

Turner, J. S. (2019). Homeostasis as a fundamental principle for a coherent theory of brains. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 374:20180373. doi: 10.1098/rstb.2018.0373

Tweedy, L., and Insall, R. H. (2020). Self-generated gradients yield exceptionally robust steering cues. *Front. Cell Dev. Biol.* 8:133. doi: 10.3389/fcell.2020.00133

Tweedy, L., Thomason, P. A., Paschke, P. I., Martin, K., Machesky, L. M., Zagnoni, M., et al. (2020). Seeing around corners: cells solve mazes and respond at a distance using attractant breakdown. *Science* 369:eaay9792. doi: 10.1126/science.aay9792

Urrios, A., Macia, J., Manzoni, R., Conde, N., Bonforti, A., de Nadal, E., et al. (2016). A synthetic multicellular memory device. *ACS Synth. Biol.* 5, 862–873. doi: 10.1021/acssynbio.5b00252

Valentini, G., Moore, D. G., Hanson, J. R., Pavlic, T. P., Pratt, S. C., and Walker, S. I. (2018). "Transfer of information in collective decisions by artificial agents," in *Proceedings of the the 2018 Conference on Artificial life: A Hybrid of the European Conference on Artificial life (ECAL) and the International Conference on the Synthesis and Simulation of Living Systems (ALIFE)*, (Cambridge, MA: MIT Press), 641–648. doi: 10.1371/journal.pone.0168876

Van Baalen, M. (2013). "The unit of adaptation, the emergence of individuality, and the loss of sovereignty," in *Vienna Ser Theor Bio*, ed. F. B. A. P. Huneman (Cambridge, MA: MIT Press), 117–140.

Vandenberg, L. N., Adams, D. S., and Levin, M. (2012). Normalized shape and location of perturbed craniofacial structures in the Xenopus tadpole reveal an innate ability to achieve correct morphology. *Dev. Dyn.* 241, 863–878. doi: 10.1002/dvdy.23770

Vergassola, M., Villermaux, E., and Shraiman, B. I. (2007). 'Infotaxis' as a strategy for searching without gradients. *Nature* 445, 406–409. doi: 10.1038/nature05464

Versteeg, E. J., Fernandes, T., Guzzo, M. M., Laberge, F., Middel, T., Ridgway, M., et al. (2021). Seasonal variation of behavior and brain size in a freshwater fish. *Ecol. Evol.* 11, 14950–14959. doi: 10.1002/ece3.8179

Vetere, G., Tran, L. M., Moberg, S., Steadman, P. E., Restivo, L., Morrison, F. G., et al. (2019). Memory formation in the absence of experience. *Nat. Neurosci.* 22, 933–940. doi: 10.1038/s41593-019-0389-0

Vine, A. L., and Bertram, J. S. (2002). Cancer chemoprevention by connexins. *Cancer Metastasis Rev.* 21, 199–216. doi: 10.1023/a:1021250624933

Vladimirov, N., and Sourjik, V. (2009). Chemotaxis: how bacteria use memory. *Biol. Chem.* 390, 1097–1104. doi: 10.1515/BC.2009.130

Volkov, A. G., Toole, S., and WaMaina, M. (2019). Electrical signal transmission in the plant-wide web. *Bioelectrochemistry* 129, 70–78. doi: 10.1016/j.bioelechem.2019.05.003

von Dassow, G., and Munro, E. (1999). Modularity in animal development and evolution: elements of a conceptual framework for EvoDevo. *J. Exp. Zool.* 285, 307–325. doi: 10.1002/(sici)1097-010x(19991215)285:4<307::aid-jez2>3.0.co;2-v

von der Ohe, C. G., Darian-Smith, C., Garner, C. C., and Heller, H. C. (2006). Ubiquitous and temperature-dependent neural plasticity in hibernators. *J. Neurosci.* 26, 10590–10598. doi: 10.1523/JNEUROSCI.2874-06.2006

Voskoboynik, A., Simon-Blecher, N., Soen, Y., Rinkevich, B., De Tomaso, A. W., Ishizuka, K. J., et al. (2007). Striving for normality: whole body regeneration through a series of abnormal generations. *FASEB J.* 21, 1335–1344. doi: 10.1096/fj.06-7337com

Wagner, G. P., Pavlicev, M., and Cheverud, J. M. (2007). The road to modularity. *Nat. Rev. Genet.* 8, 921–931. doi: 10.1038/nrg2267

Wang, X., Veruki, M. L., Bukoreshtliev, N. V., Hartveit, E., and Gerdes, H. H. (2010). Animal cells connected by nanotubes can be electrically coupled through interposed gap-junction channels. *Proc. Natl. Acad. Sci. U.S.A.* 107, 17194–17199. doi: 10.1073/pnas.1006785107

Warwick, K., Nasuto, S. J., Becerra, V. M., and Whalley, B. J. (1998). "Experiments with an in-vitro robot brain," in *Computing with Instinct. LNAI 5897*, ed. Y. Cai (Berlin: Springer).

Watson, R. A., and Szathmary, E. (2016). How can evolution learn? *Trends Ecol. Evol.* 31, 147–157. doi: 10.1016/j.tree.2015.11.009

Watson, R. A., Buckley, C. L., Mills, R., and Davies, A. (eds) (2010). "Associative memory in gene regulation networks," in *Proceedings of the Artificial Life Conference XII*, (Odense).

Watson, R. A., Mills, R., Buckley, C. L., Kouvaris, K., Jackson, A., Powers, S. T., et al. (2016). Evolutionary connectionism: algorithmic principles underlying the evolution of biological organisation in evo-devo, evo-eco and evolutionary transitions. *Evol. Biol.* 43, 553–581. doi: 10.1007/s11692-015-9358-z

Watson, R. A., Wagner, G. P., Pavlicev, M., Weinreich, D. M., and Mills, R. (2014). The evolution of phenotypic correlations and "developmental memory". *Evolution* 68, 1124–1138. doi: 10.1111/evo.12337

Wentlandt, K., Samoilova, M., Carlen, P. L., and El Beheiry, H. (2006). General anesthetics inhibit gap junction communication in cultured organotypic

hippocampal slices. *Anesth. Analg.* 102, 1692–1698. doi: 10.1213/01.ane.0000202472.41103.78

West, S. A., Fisher, R. M., Gardner, A., and Kiers, E. T. (2015). Major evolutionary transitions in individuality. *Proc. Natl. Acad. Sci. U.S.A.* 112, 10112–10119. doi: 10.1073/pnas.1421402112

Williams, K., Bischof, J., Lee, F., Miller, K., LaPalme, J., Wolfe, B., et al. (2020). Regulation of axial and head patterning during planarian regeneration by a commensal bacterium. *Mech. Dev.* 163:103614. doi: 10.1016/j.mod.2020.103614

Wolfe, C. T. (2008). Introduction: vitalism without metaphysics? Medical vitalism in the enlightenment. *Sci. Context* 21, 461–463. doi: 10.1017/s0269889708001919

Xue, Y., and Acar, M. (2018). Mechanisms for the epigenetic inheritance of stress response in single cells. *Curr. Genet.* 64, 1221–1228. doi: 10.1007/s00294-018-0849-1

Yang, C. Y., Bialecka-Fornal, M., Weatherwax, C., Larkin, J. W., Prindle, A., Liu, J., et al. (2020). Encoding membrane-potential-based memory within a microbial community. *Cell Syst.* 10, 417–423.e3. doi: 10.1016/j.cels.2020.04.002

Yang, C., Tibbitt, M. W., Basta, L., and Anseth, K. S. (2014). Mechanical memory and dosing influence stem cell fate. *Nat. Mater.* 13, 645–652. doi: 10.1038/nmat3889

Zahn, N., Levin, M., and Adams, D. S. (2017). The Zahn drawings: new illustrations of Xenopus embryo and tadpole stages for studies of craniofacial development. *Development* 144, 2708–2713. doi: 10.1242/dev.151308

Zhao, J., Yu, H., Luo, J. H., Cao, Z. W., and Li, Y. X. (2006). Hierarchical modularity of nested bow-ties in metabolic networks. *BMC Bioinformatics* 7:386. doi: 10.1186/1471-2105-7-386

Zoghi, M. (2004). Cardiac memory: do the heart and the brain remember the same? *J. Interv. Card Electrophysiol.* 11, 177–182.

Author Disclaimer: The opinions expressed in this publication are those of the author(s) and do not necessarily reflect the views of the John Templeton Foundation.

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

GLOSSARY

The following definitions of terms used in this paper (in alphabetical order) represent ways of thinking about specific terminology in the context of the proposed TAME framework. These terms have many definitions in other frameworks and are tightly interwoven, and it is likely impossible to do them full justice at this point in time (or provide uncontroversial definitions that everyone will agree capture everything of importance). Moreover, much like a theorem and its component statements, the utility of these highly-related concepts is maximized by the entire set taken together, not by crisp demarcations of any one term. The below definitions are not claimed to be uniquely correct, but merely useful; this field is still sufficiently young with respect to very basic questions, which excessively sharp definitions can limit more than they enable.

- Agency—a set of properties closely related to decision-making and adaptive action which determine the degree to which optimal ways to relate to the system (in terms of communication, prediction, and control) require progressively higher-level models specified in terms of scale of goals, stresses, capabilities, and preferences of that System as an embodied Self acting in various problem spaces. This view of agency is related to those of autopoiesis (Maturana and Varela, 1980) and anticipatory systems (Rosen, 1985).
- Consciousness—the first-person phenomenal experience of any Self—that which makes *my toothache* irreducibly different to me than anyone else's toothache or third-person descriptions of toothaches. The degree and content of consciousness is "what it is like" to be that Self, as opposed to studying it from the outside, whether or not the Self is advanced enough to be able to verbalize it or to think about it (Nagel, 1974). Consciousness here is not meant to necessarily indicate advanced, reflexive, verbal self-consciousness but rather the basal sentience (sense-process-respond loop) which is taken to be a continuum. Moreover, because all cognitive agents are inevitably made of parts, we are all collective intelligences in a strong sense (Schwitzgebel, 2015)—what it is like to be you is exactly what it's like to be a (particularly organized) collection of cells.
- Cognition—all of the activities undertaken by a Self, at whatever scale and of whatever material implementation, that underlie its gathering, processing, and acting on information for the purposes of adaptive action and perdurance against dissipation. Components include active inference, learning, and basal goal-directed activity, as well as complex cognitive skills such as symbolic reasoning, composition of concepts, language, and meta-cognition.
- Decision—an event during the traversal of some relevant space by a system's state which is efficiently modeled as a choice between diverse options. The degree of "decision-making" of any given system is proportional to the spatio-temporal and complexity distance between the events that eventually gave rise to a specific outcome and the outcome itself. Advanced Selves have inputs to their decision-making machinery that are counterfactual future states. The scale at which one defines appropriate inputs (stimuli) to a system is whatever scale is most efficient for understanding the resulting decisions (Noble, 2012; Pezzulo and Levin, 2016; Flack, 2017).
- Mind—the functional, dynamic aspect of a Self that results from all of its cognitive and somatic activities, which represents the propensities for certain types of actions and possesses some degree of sentience as a first-person perspective that perdures across changes in the material components of the body.
- Intelligence—the functional ability to solve problems in various spaces (not necessarily in 3D space), not tied to specific implementations, anatomical structures, or time scales. The degree of intelligence (IQ) is proportional to competency in navigating these spaces, including especially the ability to identify paths that temporarily lead further from the goal state but eventually enable better results. Advanced intelligence exploits additional levels of self-modeling which enables multiple levels of virtual modeling of the Self and its outside world (counterfactual thought), anxiety, and creativity (identifying opportunities, as opposed to only solving problems existing right now). In particular, by focusing on the functional aspects of intelligence, and by recognizing that there is no intelligent agent that is not made of parts, Collective Intelligence is generalized here (emphasizing the architecture of functional connections between subunits) and is not viewed as a radically distinct natural kind.
- Maslow's Hierarchy of Needs–a motivational theory of psychology that focuses on the relative types of preferences and goals which human (or other) systems pursue at various stages and scales of observation (Maslow, 1943). It also stresses degrees of integration and the modulation of higher levels by the level of stress in subunits.
- Self—a coherent system emerging within a set of integrated parts that serves as the functional owner of associations, memories, and preferences, and acts to accomplish goals in specific problem spaces where those goals belong to the collective and not to any individual sub-component. Selves are defined by the spatio-temporal scale and nature of the types of goals they can pursue—their "cognitive light cone." They have functional boundaries and material implementations but are not identical with any specific type of substrate, and can overlap within other Selves at the same, higher, and lower-level Selves. A Self is a theoretical construct posited by external systems (such as scientists, engineers, and conspecifics) and by systems themselves (*via* internal self-models), which facilitates prediction and adaptive behavior by serving as an efficient, high-level target for intervention and control strategies.
- Stress—a system-level state which serves as a driver for homeostatic loops (operating over a variable that is progressively reduced as activity gets the system closer to its desired region of action space). The spatio-temporal and complexity scale of events that can possibly stress a system are a good indicator of that system's cognitive sophistication. Stress can arise *via* discord between

external states and the Self's needs, between sensory stimuli and expectations, or between the goals of multiple subsystems within an agent, either within or across levels of organization. Thus, geometric frustration (Sadoc and Mosseri, 2007) and material scientists' notions of stress as a high-level determinant of system behavior over time (Batterman and Rice, 2014; Batterman, 2015) are minimal examples of the fundamental concept of Stress, on the same continuum as metabolic stress in bacteria, competing cellular alignment forces in planar polarity of tissues, and "true psychological stress" in organisms.

frontiers | Frontiers in Systems Neuroscience

SYSTEMATIC REVIEW
published: 12 April 2022
doi: 10.3389/fnsys.2021.787139

Cephalopod Behavior: From Neural Plasticity to Consciousness

Giovanna Ponte[1*†], *Cinzia Chiandetti*[2†], *David B. Edelman*[3,4], *Pamela Imperadore*[1], *Eleonora Maria Pieroni*[4] *and Graziano Fiorito*[1]

[1] Department of Biology and Evolution of Marine Organisms, Stazione Zoologica Anton Dohrn, Naples, Italy, [2] Department of Life Sciences, University of Trieste, Trieste, Italy, [3] Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH, United States, [4] Association for Cephalopod Research 'CephRes' a non-profit Organization, Naples, Italy

It is only in recent decades that subjective experience - or consciousness - has become a legitimate object of scientific inquiry. As such, it represents perhaps the greatest challenge facing neuroscience today. Subsumed within this challenge is the study of subjective experience in non-human animals: a particularly difficult endeavor that becomes even more so, as one crosses the great evolutionary divide between vertebrate and invertebrate phyla. Here, we explore the possibility of consciousness in one group of invertebrates: cephalopod molluscs. We believe such a review is timely, particularly considering cephalopods' impressive learning and memory abilities, rich behavioral repertoire, and the relative complexity of their nervous systems and sensory capabilities. Indeed, in some cephalopods, these abilities are so sophisticated that they are comparable to those of some higher vertebrates. Following the criteria and framework outlined for the identification of hallmarks of consciousness in non-mammalian species, here we propose that cephalopods - particularly the octopus - provide a unique test case among invertebrates for examining the properties and conditions that, at the very least, afford a basal faculty of consciousness. These include, among others: (i) discriminatory and anticipatory behaviors indicating a strong link between perception and memory recall; (ii) the presence of neural substrates representing functional analogs of thalamus and cortex; (iii) the neurophysiological dynamics resembling the functional signatures of conscious states in mammals. We highlight the current lack of evidence as well as potentially informative areas that warrant further investigation to support the view expressed here. Finally, we identify future research directions for the study of consciousness in these tantalizing animals.

Keywords: cephalopods, behavioral plasticity, cognition, consciousness, neural plasticity

OPEN ACCESS

Edited by:
Olivia Gosseries,
University of Liège, Belgium

Reviewed by:
Shuichi Shigeno,
Osaka University, Japan
Louis Neal Irwin,
The University of Texas at El Paso, United States

***Correspondence:**
Giovanna Ponte
giovanna.ponte@szn.it

†These authors have contributed equally to this work

Received: *30 September 2021*
Accepted: *22 December 2021*
Published: *12 April 2022*

Citation:
Ponte G, Chiandetti C, Edelman DB, Imperadore P, Pieroni EM and Fiorito G (2022) Cephalopod Behavior: From Neural Plasticity to Consciousness. Front. Syst. Neurosci. 15:787139. doi: 10.3389/fnsys.2021.787139

INTRODUCTION

The notion that an animal like the cephalopod mollusc *Octopus vulgaris*, an invertebrate, and its allied species (e.g., cuttlefish and squid) could have anything remotely resembling subjective experience is quite likely to be met with astonishment in some quarters. The suggestion that these animals might have a sophisticated form of consciousness would, to many, be shocking. However, from a purely theoretical perspective, the subjective experiences which we, as humans, frequently report can only lead us to *presume* that other humans have consciousness, just as we *presume* the existence of features of the external world (Humphrey, 2006; Andrews, 2020). By this line of

thinking, there is no reason to deny that many other animals experience at least some degree of primary consciousness, contingent, of course, on their sensory, cognitive, physical, and life faculties and constraints (e.g., Feinberg and Mallatt, 2020).

Consciousness has long been imagined to be very similar to what has been termed a 'first principle' in mathematics: a concept intrinsic to everyone which cannot be explained by formal logical systems because our conventional language is incapable of encompassing or expressing it. It is a concept that seems to vacillate between the domains of philosophy, science, and, sometimes, human morality (Vitti, 2010; Seager, 2016). Notably, William James was the first to objectively define consciousness not as a concept or thing, but rather as a process resulting from the complex interaction of brain, body, and environment (James, 1977). This definition helped to establish a dynamic vision of consciousness as the subjective experience of an individual that does not necessarily require explicit terms to be recognized, but rather relies on a specific, relatively complex neural organization, potentially extending this faculty to those non-human vertebrates possessing similar - or homologous - organization. Undoubtedly, having established that consciousness is a process that is endemic to the biological world, the next step might be to analyze it using Tinbergen's four questions in an effort to define its adaptive, phylogenetic, causal, and ontogenetic properties (Gutfreund, 2018).

The neuroscientist Gerald Edelman offered a compelling vision of consciousness as a process contingent on richly interconnected and reentrant neural circuits capable of integrating an extraordinary number of inputs from both external environments and internal milieus. In this view, consciousness arises through the emergence of widespread, temporally linked mappings of multimodal extrinsic and intrinsic signals, i.e., sensory binding. According to Edelman, it is likely that natural selection shaped the structures and systems that produced consciousness (e.g., thalamocortical and cortico-cortical circuitry and the limbic system, among others) through a continuous fine-tuning process over millions of years (Edelman, 2003). This becomes even clearer when we consider the adaptive value of neural systems: not a merely a set of instructions, but rather, highly selective pathways of widely distributed populations of neurons - or neuronal groups - with the ability to integrate as much information as possible and process it very quickly (see Neuronal Group Selection; e.g., Neural Darwinism, in: Edelman, 1987, 1989, 2003).

Though the vision of consciousness as one consequence of rewiring and neuronal plasticity may be hard to accept for those who have claimed consciousness as the quality that separates humans from the rest of the animals, it will be even more challenging to determine if there is enough evidence accumulated in the last decade to suggest that it is not a uniquely human faculty or even an exclusive property of vertebrates. One empirical approach to the study of consciousness is to consider behavioral abilities as indicators of consciousness; despite the absence of verbal language, there are specific solutions for investigating consciousness in non-human animals (as well as human neonates) and, in particular, demonstrating the requisite degree of neural complexity, plasticity, and behavioral flexibility for subjective experience in non-human mammals, birds, some reptiles, and even certain invertebrates (Edelman et al., 2005; Seth et al., 2005; Edelman and Seth, 2009; Vitti, 2010).

Here, we review several theories of and proxies for consciousness with a particular focus on cephalopods molluscs. We highlight the strengths, drawbacks, and lacunae when considering these animals as candidates for a distinct level or degree of consciousness. The topic has been covered previously in a series of papers (Mather, 2008, 2021a,b; Edelman and Seth, 2009; Birch et al., 2020; Feinberg and Mallatt, 2020). We will highlight the essential aspects of these works.

CEPHALOPOD CONSCIOUSNESS: A SHORT OVERVIEW OF OTHERS' CONTRIBUTIONS

Cephalopods regularly challenge our assumptions about the limits of invertebrate cognition and behavior, underline the borders of our current knowledge base, and surprise us with their unique and rich repertoire of capabilities that in some instances equal or even exceed those of certain vertebrates. Sophisticated visual and tactile learning (for review see e.g., Sanders, 1975; Marini et al., 2017), as well as a kind of spatiotemporal awareness, may indirectly support the idea that cephalopods possess a form of sensory consciousness (as suggested by Mather, 2008). Perhaps the most intriguing aspect is their almost complete lack of stereotyped behaviors or fixed action patterns. Indeed, the development of well-oriented flexible responses to changes in stimuli or environmental contexts to mention some (e.g., Sanders and Young, 1940; Maldonado, 1963b, 1965; Messenger, 1973; Sanders, 1975; Darmaillacq et al., 2004; Agin et al., 2006; Marini et al., 2017; Hanlon and Messenger, 2018) suggests that cephalopods employ domain specificity - as recently proposed by Birch et al. (2020) and overviewed by Mather (2021a,b) -, a faculty strongly associated with active brain processing (Hirschfeld and Gelman, 1994) and, as we will further mention in this paper, a theory of mind (ToM), i.e., the ability to intuit the thoughts and beliefs of others by a sort of 'mind reading' faculty that requires some neural encoding of social domains (e.g., Frith and Frith, 2006; Apperly, 2011; for cephalopods see Godfrey-Smith, 2013).

As reviewed by Edelman and Seth (2009) - and particularly at the behavioral level - consciousness has been associated with behavioral flexibility (e.g., plasticity), among other attributes. Marked learning capabilities (review in e.g., Sanders, 1975; Marini et al., 2017; Hanlon and Messenger, 2018; Gutnick et al., 2021) and interindividual differences in temperament have been observed across some cephalopod species (Mather and Anderson, 1993; Sinn et al., 2001, 2008, 2010; Sih et al., 2004a,b; Adamo et al., 2006; Scheel et al., 2017; Zoratto et al., 2018; Borrelli et al., 2020; O'Brien et al., 2021). It is noteworthy that the existence of 'personalities' in octopus and other cephalopods has been considered an «interesting manifestation of individual differences», but questioned as a demonstration of «consciousness or self-monitoring» (Mather, 2008, p. 43). However, such a spectrum of temperaments termed the 'shy-bold continuum' (Wilson et al., 1994; for cephalopods see for example:

Borrelli and Fiorito, 2008; Borrelli et al., 2020), suggests that each individual develops its own unique response to a given stimulus, forming a specific workspace and an appropriate representation of the environment (Baars, 1994; Edelman et al., 2005).

Although the large body of evidence for high-level behavioral abilities might suggest the representation of a multimodal set of perceptual and motor events in the cephalopod nervous system, this alone would not be sufficient to make the case for primary consciousness. For one thing, reported behavioral indicators may occasionally be misinterpretations - often due to the application of anthropomorphism - of what was actually observed, and as such, could skew our perspective (Gutfreund, 2019). For another, advanced cognition doesn't necessarily coincide with conscious experience, even in human beings (Vallortigara, 2017). Given these caveats, what other reliable indices or proxies can we employ to probe for consciousness in cephalopods?

It has been persuasively argued that in order for animals to evince any degree of consciousness, they must possess highly elaborated neural structures capable of generating a 'global workspace' in which sub-networks of signals (or percepts) from otherwise disparate inputs are bound together within a single dynamic network. This vast network, comprising a conscious gestalt of bound percepts, may then be broadcast widely throughout the brain, at which point complex responses can be elicited (Baars, 1994). The Global Workspace Theory (GWT) was clearly formulated with the mammalian brain in mind. GWT proposes that conscious gestalts arise specifically from signaling both within the cerebral cortex and between cortex and thalamus. The theory has been further refined through physiological and imaging data gathered primarily from human subjects (Baars, 1994, 2002; DeHaene and Changeux, 2004; Dehaene and Changeux, 2005).

Given the foregoing, is it reasonable to extend GWT to investigations of consciousness in animals quite distant from the mammalian (or even vertebrate) line?

Seth et al. (2005) proposed 14 criteria (actually 17: three well-established brain correlates + 14 distinctive properties; see also **Table 1**) for consciousness in humans and non-human mammals, including specific neuroanatomical features and physiological markers, that - with proper considerations and the right tools - could be objectively applied to non-mammalian phyla as well (Edelman et al., 2005). Extending this framework to cephalopods, a logical starting point would be the identification of reliable objective neural correlates analogous to those observed in conscious, awake vertebrates.

At the level of gross morphology, invertebrates such as cephalopods lack neural structures resembling cortex or thalamus, but this does not preclude the existence of structures that carry out functions closely analogous to those of cortex and thalamus (Shigeno et al., 2018).

Indeed, the architectural complexity of certain neural structures in cephalopods approaches that of higher vertebrates, and the specific organization of those structures suggests they may be functional analogs of mammalian brain areas implicated in the instantiation of conscious states. In particular, the dorsal basal and subvertical lobes receive disparate inputs from throughout the body *via* both direct and indirect pathways from the suboesophageal mass, which acts as a relay center - akin to the thalamus—conveying signals to the frontal and vertical lobes. Intriguingly, the purpose of these lobes is to integrate such inputs and produce an elaborate response: a function similar to that performed by the mammalian cortex (see Shigeno et al., 2018 and below).

According to Carls-Diamante, despite the outstanding cognitive abilities that might underlie a sort of 'mentality,' cephalopod consciousness may not be organized as a 'united field' (Carls-Diamante, 2017, 2021), but rather as a "community of minds" (Schwartz, 2019). The assumption of Carls-Diamante derives from the so called 'isomorphism thesis,' which relates the structure of an animal's consciousness to the structure of its nervous system (Bayne, 2010).

The atypical neuroanatomy of cephalopods, composed of a central brain and a complex peripheral nervous system - comprising about two thirds of the total number of neural cells (in the octopus; e.g., Young, 1963; Graziadei, 1971) supports the view of an embodied organization of the octopus nervous system (Hochner, 2013). Such independence from the brain mass (i.e., the neuro-motor system in the arms; Sumbre et al., 2001, 2005, 2006) could arguably be sufficient to rule out the possibility that cephalopods integrate different experiences into a perceptual unity (Carls-Diamante, 2017). However, the structure of subjective experience conceived as a united conscious field is too simplistic, and it can hardly be generalized to the organization of many species' sensorimotor systems, as these are very different from one another (van Woerkum, 2020). Thus, with these assumptions in mind, what cephalopod experience can only be understood by investigating how their body, nervous system, and sensory equipment allow them to actively and flexibly interact with - and, integrate inputs from - the environment (Godfrey-Smith, 2019; van Woerkum, 2020; see also Masciari and Carruthers, 2021).

In a recent attempt to address the issue of consciousness in non-mammal animals, Birch et al. (2020) suggested a framework comprising five dimensions that draws from examples provided by the octopus and other closely allied invertebrate candidates. In this synthesis, the five dimensions are: perceptual richness (p-richness), evaluative richness (e-richness), integration both at a given time (unity) and across time (temporality), and self-awareness (e.g., selfhood). Notably, a discussion around p-richness in octopus has recently been initiated by Mather (2021b). **Table 1** lays out a possible correspondences between these five dimensions and the criteria specified by Seth et al. (2005).

In what follows, we will further explore these issues and argue that if subjective experience requires some minima of network organization and computational power and primary consciousness can be imputed from simple sensory and cognitive representations, then a cephalopod like the octopus - with its 500 million neurons - large and highly differentiated central nervous system, and rich behavioral repertoire and flexibility, must be considered as a fruitful model organism for investigating consciousness beyond the vertebrate lineage. And so begins our journey to the biological frontiers of awareness on the backs of the cephalopod molluscs.

TABLE 1 | Key features to access consciousness dimensions, and dimensions and hallmarks of consciousness (modified after: Edelman et al., 2005; Seth et al., 2005; Edelman and Seth, 2009; Birch et al., 2020).

Key features for assessing dimensions of consciousness		
	Features	**Notes**
	EEG signatures	Evidence in cephalopods: Fast, irregular electrical brain activity; compound field potentials, and evoked potentials (Bullock, 1984; Bullock and Budelmann, 1991; Brown et al., 2006; see also Butler-Struben et al., 2018)
	Cortex and thalamus	Evidence in cephalopods: existence of functional analogs identified at the level of the superior frontal-vertical lobe systems and dorsal basal lobe (supra-esophageal mass; for review see Shigeno et al., 2018)
	Widespread brain activity	For cephalopods see: Bullock (1984), Bullock and Budelmann (1991), Brown et al. (2006), Butler-Struben et al. (2018)
Wide Range Dimensions	**Hallmarks**	**Notes**
p-richness	Sensory binding	Ability to perceive different features of the environment (e.g., shape, taste, odor). For review in relationship to cephalopods, see Birch et al. (2020), and Mather (2021b); see also e.g., Chiao and Hanlon (2001a,b), Scatà et al. (2017), Mezrai et al. (2019), van Giesen et al. (2020).
	Facilitation of learning	Conscious perception and learning of temporal relationships. As reviewed by Birch et al. (2020), and also Mather (2021b) for cephalopods; see also, e.g., Marini et al. (2017), Borrelli et al. (2020), Bublitz et al. (2017) Schnell et al. (2021a). In regards to neural correlates of sensory and learning processing/capacities in cephalopods, note the large number of studies based on the impairment of the neural circuitry underlying visual and chemo-tactile memory-systems (for review see: Sanders, 1975; Borrelli and Fiorito, 2008; Marini et al., 2017).
e-richness	Accurate reportability	For review in relation to cephalopods, see Birch et al. (2020). Conscious contents are reportable by several behavioral responses following evaluation. Such valence should be applied in affectively based decision making.
	Informativeness	Some animals may continually evaluate small changes in their internal milieus and external environments, while others may only react to substantial changes and ignore redundant stimuli. For review, see also: Marini et al. (2017), Hanlon and Messenger (2018)
	Focus-fringe structure	"Fringe Conscious" (e.g., Norman, 2017) events, like feelings of familiarity and experiences having emotional valence (e.g., the "tip of the tongue phenomenon") are consistent with the idea that the self is an interpreter of conscious experience rather than a primary source of perceptual content.
Unity	Subjectivity	For review in relation to cephalopods, see Birch et al. (2020) and also Mather (2021a). Consciousness is marked by the existence of a private flow of events accessible only to the experiencing subject, despite being reportable. The world and all the experiences generated by our brain have a common subject.
	Internal consistency	Consciousness is marked by a consistency constraint. That is, even when similar stimuli are presented simultaneously, only one can become conscious at any given time (in relation to our "inner" common subject).
	Limited capacity and seriality	Consciousness flows from one scene to another in a serial manner and is constrained to just one scene at any given moment.
	Self-attribution	Consciousness is experienced by an observing self. As reviewed by Birch et al. (2020), behavioral and neural asymmetries have been reported in cephalopods to different extents (Jozet-Alves et al., 2012a; Schnell et al., 2016a; Frasnelli et al., 2019). Cephalopod molluscs are capable of parallel processing of visual and/or tactile inputs (e.g., Borrelli, 2007; Schnell et al., 2018; Borrelli et al., 2020), possibly including recognition (e.g., Tricarico et al., 2011, 2014; Nesher et al., 2014; Katz et al., 2021), but also in the process of integrating information from both eyes (e.g., Feord et al., 2020; see also El Nagar et al., 2021).
Temporality	The rapidly adaptive and fleeting nature of conscious scenes	Birch et al. (2020) mention some cephalopod studies that may address this dimension (e.g., Billard et al., 2020a,b; Poncet et al., 2020; Schnell et al., 2021a,b). Immediate experience of sensory past and cognitive present that persists for a few seconds, forming a continuous stream of events.
	Stability of contents	Conscious contents are stable, even though experiences can be temporally integrated across longer timescales—comprising past and future events—in what it might be termed "temporal dimensions."
Selfhood	Conscious knowing and decision making	Consider here: Crook and Walters (2014), Birch et al. (2020); and for example (Tricarico et al., 2011, 2014; Nesher et al., 2014). Consciousness is useful for generating knowledge about the world around us, as well as that of our own internal states and processes, all of which can influence decision making.
	Allocentricity	The foregoing implies the discrimination of ourselves from the external world by an allocentric faculty which makes use of neural representations of external objects to build conscious scenes.

Wide range: Consciousness has an extraordinary range of different contents, including perception in the various senses, endogenous imagery, feeling states, inner speech, concepts, action-related ideas, and "fringe" experiences such as feelings of familiarity. Examples of studies testing a given feature, - suggesting the possibility of a given hallmark in cephalopods, within a specific dimension are based on the recent review by Birch et al. (2020) and our own coverage of the scientific literature.

CEPHALOPOD COGNITION

Prolog: A Narrative Arc From Aristotle and Darwin to the Present

Throughout the history of natural observation and exploration, cuttlefish, squid, and octopuses (the most thoroughly studied of all cephalopods) have provided ample and compelling evidence that they are more than simply eight-armed (plus two tentacles in the case of cuttlefish and squid) masses of muscles guided by a primal insatiable appetites.

For many observers through the ages, these animals have offered ample cause for surprise, exhibiting behaviors that ultimately contributed to their steady rise in public awareness over the past century (Lee, 1875; Lane, 1960; Cousteau and Diolé, 1973; Mather et al., 2010; Courage, 2013; Schweid, 2013; Montgomery, 2015).

Apart from Aristotle - who in the fourth century BC expressed the opinion that the octopus was a curious, but stupid animal (Aristotle, 1910; overview in Marini et al., 2017) - recorded observations of, and anecdotes about cephalopods and their astonishing capabilities are ubiquitous throughout recorded history (for review see for example, Borrelli and Fiorito, 2008; De Sio, 2011; Dröscher, 2016; Marini et al., 2017). Interestingly in *The Voyage of the Beagle*, Darwin reported an encounter with an octopus on the coast of Cape Verde. He noted that the octopus was not only able to withstand his gaze, but also seemed to stare back at him intently in a kind of match of attentional wits (Darwin, 1870).

Octopuses were depicted in ancient bestiaries as voracious, cunning, and positively evil (see for example: Lee, 1875; Chapko et al., 1962), based on the broad attribution of moral categories to animals. Setting aside this compellingly colorful characterization, a considerable number of octopus tales, including Darwin's, have, over the centuries, contributed to cementing the reputation of the octopus in Western culture (Lane, 1960; Caillois, 1973; Hochner et al., 2006; Borrelli and Fiorito, 2008; Makalic, 2010; Hochner, 2012; Marini et al., 2017).

In the first century AD, Pliny the Elder (Pliny, 1961) reported witnessing an octopus waiting patiently for a large shellfish (*Pinna nobilis*) to open its valves in order to prop it open with a stone it held in its arms. The veracity of Pliny's observation was corroborated centuries later by the personal account of the zoologist Jeannette Powers (Power, 1857; for an overview see: Borrelli and Fiorito, 2008; Marini et al., 2017).

With the advent of marine stations and inland aquaria, cephalopods became more readily amenable to observation and experimentation, and a wealth of comparable histories emerged, based on more or less systematic and increasingly frequent observations (De Sio et al., 2020). For example, there are anecdotal accounts of cephalopods that recognized individual conspecifics, performing body patterns to communicate, exhibited individual temperaments, and even recognized individual humans (Romanes, 1885) possibly forming a kind of affective bond.

Following from such compelling anecdotal accounts, experimental trials have provided proof of such sophisticated capabilities in cephalopods. Among the most notable findings are: the recognition of human faces (Anderson et al., 2010) and individuals (Boal, 2006; Tricarico et al., 2011, 2014), play (Mather and Anderson, 1999; Kuba et al., 2003, 2006), 'personality' (Mather and Anderson, 1993; Borrelli, 2007; Borrelli and Fiorito, 2008; Sinn et al., 2008, 2010; Borrelli et al., 2020; O'Brien et al., 2021), social learning (Fiorito and Scotto, 1992; Fiorito, 1993; Fiorito and Chichery, 1995; Amodio and Fiorito, 2013; Huang and Chiao, 2013; Tomita and Aoki, 2014), episodic memory (e.g., Pronk et al., 2010; Jozet-Alves et al., 2013; Schnell et al., 2021b), and deliberate and projective tool use within a specific octopus population (Finn et al., 2009), among others.

Discriminatory and Anticipatory Behaviors

A strong link between perception and memory, together with the functional neural circuitry underlying such a link, have been proposed as necessary requisites for conscious processing (Edelman, 1993; Edelman et al., 2011). Since discriminatory and anticipatory behaviors suggest such a link, we will now turn our focus to evidence for these behaviors in cephalopods.

As reviewed extensively by Marini et al. (2017), the breadth of learning paradigms in which cephalopods have demonstrated their cognitive capabilities is quite remarkable (see Table 3 in Marini et al., 2017; for review see also Hanlon and Messenger, 2018). Indeed, the vast majority of the learning studies carried out on cephalopods have relied on their well-characterized predatory behavior. Such behavior has been leveraged both to study the recovery of predatory performance following capture and to evaluate the possible interference of various stimuli or contexts with the animals' attack response. As a consequence, an established practice of learning paradigms for octopuses and other cephalopods (mainly *Sepia officinalis*) is that given experimental protocol should start after a period of 'acclimatization' (sensu Boycott, 1954; Maldonado, 1963b,c) for an animal in a captive situation. It is since the pioneering studies initiated at the end of the 1940's up to recent times, the acclimatization period is a variable length of time during which the animal is exposed to a novel environment (e.g., the tank and its surroundings) and presented with a live prey or conditioned to attack dead prey (Amodio et al., 2014; Fiorito et al., 2015). The 'acclimatization' (i.e., acclimation) is considered a form of contextual learning (Maldonado, 1963a,b,c, 1965; Borrelli, 2007; Marini et al., 2017; Borrelli et al., 2020). Trials with cuttlefish, octopus, and in some cases squid, have shown that the daily presentation of food increases the likelihood that the animal will attack. Predatory performance, measured as the time to attack prey from its appearance in the tank, thus improves over time. This phenomenon reveals an important feature of both 'positive' and 'negative' learning capabilities (Maldonado, 1963b, 1965). Notably, this process is regulated mainly by the vertical lobe system, as shown in octopus (Maldonado, 1963a; review in Sanders, 1975). Also evident from these studies are the differences between individuals; as contextual learning progresses, other characteristics of the subjects may become apparent, with inter-individual differences emerging in response (e.g., readiness) to stimuli (Borrelli, 2007; Borrelli et al., 2020). Of course, it should be noted that differences between species are to be expected,

owing to different lifestyles and adaptive capabilities (Nixon and Young, 2003; Hanlon and Messenger, 2018; Ponte et al., 2021).

Moreover, the exposure to a novel laboratory environment - e.g., the tank and/or experimental setting - involves the confinement of a given animal to a space that comprises a much smaller foraging area than that encountered in the wild. The animal is thus immersed in a comparatively monotonous captive setting, regardless of the degree of enrichment provided. Under these circumstances, evidence of the extreme breadth of cephalopod behavioral plasticity again comes to the fore. Frequently, animals have been presented with tasks - even those spaced across different trials - designed to assess their predatory behavior (e.g., attack/non-attack or take/reject responses). In such cases, they have generally adapted their species-specific predatory response to the new context. This type of contextual learning takes a variable amount of time and depends on the species being investigated, the animals' previous experiences, individual variability due to ecological and biological factors, including developmental and life cycle stages (e.g., differences in age, sex, maturity, etc.), neophobia, interindividual variability in behavioral responses (e.g., temperament), and plasticity (Borrelli et al., 2020), among others. Such studies of predatory behavior once again provide clear examples of a positive learning process (Maldonado, 1963b, 1965).

Individual and social learning have been widely explored in cephalopods (Sanders, 1975; Marini et al., 2017; Hanlon and Messenger, 2018) and, as previously noted, in all cases animals have exhibited a high degrees of plasticity and adaptability in their behavior. A few examples bear mentioning here.

Addition of quinine (a bitter taste substance) to the carapace of presented prey, such as crab or shrimp, resulted in rapid learning of taste aversion in *Sepia officinalis*; this facilitated the animals' future choice of prey and behavioral responses were retained over long durations (Darmaillacq et al., 2004). Images of a potential predator (e.g., a bird) gliding over the tank elicited startle reactions in cuttlefish (Calvé, 2005), which also affected future hunting behavior (Adamo et al., 2006). Successive visual discrimination tasks have generally been used as training protocols for octopus (for review see for example: Sanders, 1975) as well as cuttlefish in which autoshaping has been demonstrated (Cole and Adamo, 2005). The foregoing provide further examples of the classic training paradigm that has been well-established for cuttlefish (i.e., the "prawn-in-the-tube;" e.g., Sanders and Young, 1940; Messenger, 1973; Agin et al., 2006; Purdy et al., 2006; Cartron et al., 2013) and the visual (and tactile) discrimination tasks or problem-solving paradigms that have been developed for octopuses over many years (Sanders, 1975; Wells, 1978; Marini et al., 2017).

Memory retrieval is the fundamental basis for an individual's ability to benefit from past experiences. In some cases, though, it proves particularly useful in referencing a specific episode and where and when that episode occurred. In addition to studies in birds and mammals, recent work in cuttlefish has shown that these animals remember what they ate, as well as where and how long ago they ate, thus satisfying the "what," "where," and "when" criteria for episodic-like memory (Crystal, 2010; Jozet-Alves et al., 2013; see also e.g., Schnell et al., 2021b). Furthermore, while episodic memory refers, in a sense, to the ability to time-'travel' to an individual's past, retrieving specific features belonging to such memories is a cognitive capacity that involves the contextual activation of source-memory processes. In other words, there must necessarily be semantic processes in play that afford the retrieval of a memory and its origin, as well as an indexical comparison with other stored information in order to distinguish different episodic memories from one another. Studies of *S. officinalis* proved these animals' ability to discriminate between visual and olfactory modalities and then recall which one was previously encountered before an extended delay (Billard et al., 2020a).

Notably, John Zachary Young (JZ) failed to establish definitive evidence in support of the integration of sensory modalities in octopus during learning (e.g., visual vs chemo-tactile; Young, 1991, 1995), despite extensive overlap in the neural circuitry mediating the two sensory-motor learning and memory systems, as well as some behavioral indications (Marini et al., 2017). In this regard, it would be useful to refer to the matrix-like functional organization of cephalopod nervous systems. Multiple matrices occur in regions of the central nervous systems of cephalopods. These control behavioral responses, allowing signals of different types (e.g., visual, chemo-tactile) to interact to some degree and regulate subsequent behavior - in particular, the attack/take and retreat/reject responses (Young, 1961, 1964; Maldonado, 1963c; Packard, 1963). These systems of matrices work by modulating promotion and/or inhibition of specific responses. Overall, they are tuned to facilitate the exploratory behavior that characterizes these animals. According to Young (and as emphasized on many occasions in this review), cephalopod matrix systems bear more than a passing resemblance to regions of the mammalian nervous system, in particular the limbic lobe and neocortex (Young, 1995). Despite some indications of more extensive comingling of signals, Young concluded that complete integration - or transfer - between the visual and tactile information matrices occurred only at the effector level. However, in preliminary experiments, Allen et al. (1986) showed that limited cross-modality does indeed occur (but see also Anderson and Mather, 2010). Given the limited nature of recent evidence from cuttlefish and the need for further studies of octopus sensory integration, we can only conclude that cross-modality is a long-standing issue in cephalopod cognition which warrants further systematic study. Nevertheless, the degree of behavioral complexity shown by cephalopods provides a strong rationale for exploring the possibility of conscious experience in these invertebrates.

Cephalopods commonly adopt dynamic, flexible predatory strategies that include selective, opportunistic, and plastic foraging behaviors in response to changing environmental conditions (Hanlon and Messenger, 2018). Apropos of such flexibility, recent lines of research have addressed whether cephalopods exhibit future-oriented behaviors or are capable of planning. By definition, future-oriented planning in animals (e.g., Clayton et al., 2003) requires behavior to be flexible and dependent on, or sensitive to, consequences. In cuttlefish, Billard et al. (2020b) have shown that animals adapt their behavior to environmental conditions on a daily basis. Moreover, a certain food choice in one moment of the day can determine the

dietary choice in a following moment, increasing variation (an instance of food devaluation). To rule out the possibility that an animal's future planning depends on relative or contingent motivational states, other experiments have been carried out, showing that preference of *S. officinalis* for a shrimp in a quantity comparison test occurs *via* learned evaluation that depends on the relative value of previous prey choices (Kuo and Chiao, 2020). In addition, self-control and tolerance of delays in receiving a reward - which are well-characterized features of mammals with elaborate inhibitory neural circuitry - are also documented in this species (Schnell et al., 2021a). Finally, we can expect other paradigms (e.g., maze learning) and model organisms (e.g., octopus) to shed further light on cephalopods' capacity for future-oriented behaviors and planning (Poncet et al., 2020).

The foregoing behavioral paradigms and tests of cephalopod capabilities are largely based on their predatory response, an aptitude which relies mainly on visual cues, as well as chemo-tactile information (Yarnall, 1969; for review see: Sanders, 1975; Borrelli and Fiorito, 2008; Marini et al., 2017; Villanueva et al., 2017; Hanlon and Messenger, 2018; see also Maselli et al., 2020).

As visibility in water may often be limited, chemical cues can provide alternative reliable signals that aquatic animals can use, even for the identification of conspecifics. This is the case for cephalopods (Huffard and Bartick, 2015; Polese et al., 2015; Morse et al., 2017; Morse and Huffard, 2019) as well as a wide variety of other invertebrate and vertebrate phyla, including crustaceans, insects, and fish (Hepper, 1986; Cannicci et al., 2002; Gherardi et al., 2010, 2012; Sheehan and Tibbetts, 2011). Moreover, the sense of touch, which may be linked to taste, plays an important role in octopus foraging and learning (Chase and Wells, 1986; Mather and O'Dor, 1991; Forsythe and Hanlon, 1997; Godfrey-Smith and Lawrence, 2012; van Giesen et al., 2020) and is also involved in some social interactions (Huffard et al., 2008; Amodio and Fiorito, 2013; Caldwell et al., 2015; Huffard and Bartick, 2015; Scheel et al., 2016; Morse et al., 2017).

For example, Tricarico et al. (2011) found that octopuses performed a higher number of physical contacts when placed in an arena with conspecifics they had never encountered before the testing condition, in contrast to those that had previous experience with the 'dear enemy' on the other side of a transparent barrier during a preliminary acclimation phase of the experiment (Tricarico et al., 2011).

As highly developed as cephalopods chemo-tactile faculties may be, it is their visual faculties that stand out among the invertebrates, rivaling even those of some higher vertebrates (e.g., Packard, 1972; Hanlon and Messenger, 2018). Complex vision allows animals to negotiate a wide variety of ecological and biological challenges, including predation, navigation, discrimination learning, some forms of proprioception (Wells, 1960; Gutnick et al., 2011), and even intraspecific communication. A graded diversity of color, texture and postural components forms the basis for the body patterns emitted over longer or shorter periods of time (Packard and Sanders, 1971; Packard and Hochberg, 1977; review in: Borrelli et al., 2006; Hanlon and Messenger, 2018). While body patterning allows effective mimicry and disguise, among other functions, it also provides a channel for intraspecific communication (Hanlon et al., 1999; Shashar et al., 2004; Schnell et al., 2016b), including hidden or 'secret' signals to other species (e.g., Mäthger and Hanlon, 2006; review in e.g., Tricarico et al., 2014; Hanlon and Messenger, 2018).

In many species, including a wide variety of primates and birds, vision is the primary sense used to distinguish individuals by specific facial attributes, recognize emotions by body posture and facial expression, and control gaze direction; a faculty often exploited by researchers in investigations of ToM in non-human animals (Bugnyar et al., 2004; Carter et al., 2008; Wilkinson et al., 2010; Grossmann, 2017; Nawroth et al., 2017; Kano et al., 2018). Indeed, a seemingly complex test termed "reading the mind in the eyes" has been devised to study how an adult human is able to assign a complex mental state to another simply by looking into his eyes (Baron-Cohen et al., 2001). A cursory survey of the comparative literature (see above) suggests that this aspect of 'mind-reading' may have its earliest antecedent in the sensitivity of gaze direction.

Certainly, the use of vision to recognize individuals has an ancient origin. Notably, insects such as the social wasp *Polistes fuscatus* (Tibbetts, 2002) and crustaceans such as lobsters (Gherardi et al., 2010), crayfish (Van der Velden et al., 2008) and crabs (Cannicci et al., 2002) are able to identify individual conspecifics based on unique visual facial cues.

Despite our limited understanding of social (and individual) recognition in octopuses and other cephalopods (Boal, 2006; Tricarico et al., 2011, 2014), accumulating evidence suggests the existence of a complex vision-based modality that mediates interactions between individuals (see above; Packard and Sanders, 1971; Kayes, 1974; Tricarico et al., 2011; Scheel et al., 2016; Schnell et al., 2016b). Neighbors typically show few agonistic interactions with each other, suggesting that they are affected by the "dear enemy phenomenon," i.e., a reduced aggressiveness toward neighbors in territorial animals (*sensu* Fisher, 1954), a phenomenon that has also been observed in birds, mammals, and many other vertebrates, as well as in a number of invertebrates (Tibbetts and Dale, 2007; Snijders and Naguib, 2017). The finding by Tricarico et al. (2011) that *O. vulgaris* can recognize conspecifics, discriminate known from unknown individuals, and remember the discrimination, indicates that this species is capable of at least class or binary individual recognition (Tibbetts and Dale, 2007), an ability not yet demonstrated in other cephalopod species. The ability to recognize and remember 'opponents' and conspecifics may be of adaptive value to octopuses, as it is likely the proximate mechanism regulating the 'dear enemy' phenomenon. This may explain the rare interactions between octopuses observed in the field (Tricarico et al., 2011). Despite the need for more in-depth studies to determine whether these animals are capable of true individual recognition, the work of Tricarico et al. provides to the best of our knowledge, the only known account of conspecific social recognition for this taxon (Tricarico et al., 2011, 2014). This study is comparable to the brief report by Anderson et al. (2010) that octopuses are capable of recognizing individual caretakers in the laboratory, confirming the ancient anecdote about

cephalopods mentioned earlier (Schneider, 1880; Romanes, 1885).

Some animal species have been reported to be able to differentiate among humans by their faces; an ability not limited to domesticated species (Boivin et al., 1997; Tanida and Nagano, 1998; Rybarczyk et al., 2001; Racca et al., 2010; Stone, 2010; Nagasawa et al., 2011; Müller et al., 2015; Wood and Wood, 2015), but also observed in invertebrates such as the honeybee and other insects (Dyer et al., 2005; Avarguès-Weber et al., 2017). If recognition of individual humans is confirmed in the octopus, this could provide further evidence of the cognitive distinctiveness of these animals among invertebrates.

YOUNG'S CEPHALOPOD MODEL OF THE BRAIN AND THE SEARCH FOR CONSCIOUSNESS IN CEPHALOPODS

Young (1954) famously proposed the octopus brain as a useful general model for the study of learning and memory, as it was considered both a tractable object for experimental study (e.g., a nervous system much simpler than our own) and a pointedly epistemic - even rhetorical - device that enjoins the researcher to find novel ways to discuss physical phenomena. In his view, these ways departed from our tendency as humans to frame everything in psychological terms when speaking about ourselves. In this sense, JZ claimed to be following the example of Ryle (1949). Of course, there is a vast semantic and epistemological chasm between the search for memory in a mollusc (even a cephalopod) and assessment of its higher cognitive faculties, from mind-reading to consciousness. First of all, the very concept of 'memory' needed to be progressively re(de)fined by Young to afford comparisons across the animal kingdom, as well as with genetically and electronically based memory systems (see De Sio, 2011 and cited works therein). Understandably, such a perspective might surprise the contemporary reader. We are accustomed to speaking - almost reflexively so - of the memory of a computer, the memory of our immune system, or the memory stored in our genes, often without acknowledging the analogical heavy lifting involved in drawing this seemingly simplistic equivalence. In the search for memory and its engram, a link with mechanical causality was attempted for analytical purposes by Young using octopus as the biological platform (Young, 1951).

Young considered the octopus as the animal possessing a brain appearing the «most divergent from that of mammals that is really suitable for study of the learning process» (Young, 1971, p. vii). The phenomenological proximity of behavioral traits to, and vast phylogenetic distance from, vertebrates convinced Young and his many collaborators to consider cephalopods, especially *O. vulgaris*, to be suitable general model of the brain (Young, 1964). As reviewed in Marini et al. (2017), Young and Boycott began their explorations of *O. vulgaris* at the Stazione Zoologica (Naples, Italy) in the spring of 1947, starting from scratch and based on scarce and largely unsystematic precedents. Their aim was to study learning in these animals by combining behavioral observations and surgical ablation of selected parts of the neural centers, in order to explore and define the higher functional organization of the octopus' brain and its control of behavioral outcomes. After many years of intensive study and systematic experimentation, Boycott presented an efficient - and simplified - training technique in which all the possible outcomes (e.g., complexity, individuality, and ambiguity of behavioral responses) were not considered. The predatory response of *O. vulgaris* (as in other cephalopods, e.g., Sanders and Young, 1940; Messenger, 1973) was exploited as a bio-behavioral key for teaching animals to discriminate between positively and/or negatively reinforced stimuli; in other words, to make choices and decisions in a given situation. Tens of trials were sufficient for the animals to learn the task and respond correctly in a fairly stable and predictable manner (review in Marini et al., 2017). Training animals to discriminate between different shapes by simultaneous and/or successive presentation of two discriminanda, proved successful (Sanders, 1975). The original training protocol was refined several times until it was finally standardized, such that octopuses: (i) are given a period of acclimation in the tanks; (ii) after acclimation, wait in their den until a stimulus enters their tank and elicits a response; and (iii) after a short period of attention, that response triggers either 'retreat' or 'attack' (Maldonado, 1963b,c, 1965; Packard, 1963). Of course, 'attack' and 'retreat' are not the only elements of the story; decreasing time needed to respond to the stimulus, 'attention', 'cautiousness', 'incomplete-attacks', 'shyness', and 'boldness' represent the complex behavioral responses exhibited by an octopus during training (Maldonado, 1963b, 1965; Borrelli, 2007; Borrelli et al., 2020). The versatility of the octopus training paradigms fostered the growth of the field well into the 1960's, with several directions of research evolving from the original work (see Figure 1 in Marini et al., 2017). These studies demonstrated that the octopus was capable of a diversity of learning capabilities (Sanders, 1975; Marini et al., 2017). Through this work, Young's goal was achieved: he was able to build a detailed model of the brain of a learning and behaving octopus (Young, 1961, 1964). Young was inspired by the 'proto-cybernetic' theory of learning and memory dating back to the early 1940's (Craik, 1967) and designed around a feedback process. The model was applied on several occasions (Clymer, 1973; Myers, 1992), including Clymer's application of the 'mnemon' concept in which the characterization of a visual feature induces an associated memory value resulting from experience. This leads to a system where a given visual input induces a response in a specific set of classifying cells that generates a command to attack (i.e., a predatory response) and is further summated to produce an attack 'strength' (*sensu* Maldonado, 1963c). Conversely, a retreat command is generated by opposing inputs, and their relative strengths are combined to determine the final attack/retreat response (Clymer, 1973). Similarly, another cybernetic circuit was created by Myers (1992) based on octopus' mnemon and neural networks (review in Marini et al., 2017).

As cephalopods are quite distant from, and not as well-characterized as the more familiar, systematically investigated vertebrate models for cognition, exploring the possibility of consciousness in this group of animals necessarily prompts us to

ask: how much are they really like the higher vertebrates and to what extent can they be described in terms we readily apply to ourselves, rather than *via* a more mechanistic account?

An insight by Premack and Woodruff is particularly salient here:

«As to the mental states the chimpanzee may infer, consider those inferred by our own species, for example, *purpose or intention*, as well as *knowledge*, *belief*, *thinking*, *doubt*, *guessing*, *pretending*, *liking*, and so forth» (Premack and Woodruff, 1978, p. 515).

Premack (1988) enumerated this consideration more explicitly by clarifying that the initial question for Woodruff and himself was whether apes "*do what humans do*" and therefore if attributing states of mind to individuals of another species might enable us to predict and explain the behavior of that species (Premack, 1988; see also Emery and Clayton, 2009). The fact that we have only a vague idea of how animals communicate adds another twist to the critical Kantian test (Griffin, 1976). A further complication is that inferences about the possibility that an animal has consciousness inevitably direct or (more often) indirect inferences about our relation to that animal.

In addition, confronting the mental status of a particular animal may summon consideration of the moral status of that animal, its suitability as a model for cognition and behavior, and the acceptability of its use as a commodity or as an experimental substitute for a more ethically 'indispensable' organism. This last argument is now particularly relevant for cephalopods, considering their inclusion in EU Directive 2010/63 (Smith et al., 2013; Fiorito et al., 2014; Di Cristina et al., 2015) on the grounds of the public perception of these animals and their presumed ability to feel pain and suffering (EFSA Panel, 2005; Smith et al., 2013; Di Cristina et al., 2015). The fact that they are considered to have a degree of sentience (Birch et al., 2021) and may well be capable of at least some form of sensory consciousness would likely be a step forward in defining the parameters of future cephalopod research, including, but not limited to, investigations of behavior and cognition.

Big-Brained Invertebrates That Engage With a Temporally and Spatially Variable Environment

In a popular essay published more than 30 years ago, Allan Wilson suggested that in vertebrates, there is «…an autocatalytic process mediated by the brain: the bigger the brain, the greater the power of the species to evolve biologically» (Wilson, 1985, p. 157). Taking into account increases in genome size, relative brain size, and the number and complexity of neural cell types, Wilson argued that accelerated rates of morphological change in vertebrates over the course of evolution reflected a trend toward increasing complexity of behavioral abilities. In other words, species that had evolved higher numbers and a greater diversity of brain cells and connections were also those which had undergone an increased degree of organization and elaboration of behavioral repertoires. These species were thus able to cope better with environmental changes, accelerating their evolution by adapting more quickly than species in which a lower degree of complexity was achieved. He also suggested that: «…culturally driven evolution is by no means confined to humans. Imitative learning occurs in many species having brains that are relatively large in relation to body size …. [It] may also occur in some fishes, squids and insects, although it has not yet been demonstrated in them.» (Wilson, 1985, p. 156).

Testing Wilson's Behavioral Drive Hypothesis in cephalopods remains an attractive and intriguing idea (Borrelli, 2007). Such an approach should necessarily incorporate the relationship between the nervous system and the ecology in which it is embedded (e.g., environment and lifestyle/habits; Ponte et al., 2021). It should also consider the computational capacity of the brain - not simply its size. An increase in computational power may have occurred during cephalopod evolution, considering, for example, the reduction in the size of nerve cells between the appearance of squids (i.e., the giant axon) and the later emergence of octopuses (e.g., Young, 1963; Nixon and Young, 2003). Octopus lifestyle must also be taken into account, for example the fact that most species are solitary-living and thus considered to be asocial. Of course, this is not the case for all octopus or cuttlefish species, nor does it appear to be generally true of squid species. Nevertheless, most species of cephalopods have historically been considered asocial animals in the sense that they don't establish or maintain familial relationships and are relatively short-lived (in contrast to the social mammals). Still, this overarching generalization has often been contradicted by both observation and experimental studies (Fiorito and Scotto, 1992; Fiorito, 1993; Huang and Chiao, 2013; Tomita and Aoki, 2014) as well as recent accounts (e.g., Godfrey-Smith and Lawrence, 2012; Amodio and Fiorito, 2013; Guerra et al., 2014; Scheel et al., 2016).

The foregoing prompts some important questions. As a largely asocial animal, why would the octopus have developed the capacity to learn from conspecifics (Fiorito and Scotto, 1992)? Why would a specific population of octopuses travel a fairly significant distance in order to procure coconuts to use as nests (Finn et al., 2009)? And finally, why do the actions of these animals appear intentional to such an extent that their interpretation can confound even the most experienced and least anthropocentric of observers? To this last question, a light rejoinder may have been provided by Buytendijk (1933), who stated that this putative intentionality is conveyed because octopuses give the impression of staring back - or looking you directly in the eye - such that they readily seem to cast their spell on the behavioral scientist.

IDENTIFYING POSSIBLE NEURAL SUBSTRATES FOR CONSCIOUSNESS IN CEPHALOPODS

The central brains of cephalopods have unusual features that distinguish them from the nervous systems of other molluscs (see review by Ponte et al., 2021). Among these, the most relevant are:

i. The highest degree of centralization among invertebrates (insects excluded), partly due to the shortening of connectives.

ii. The compact size of neurons acting as local interneurons (e.g., nuclear diameters of 3–5 μm), allowing for a relatively greater cell density.
iii. The reported absence of somatotopy in these animals, except in the chromatophore lobes, and tract-level representations of the labial nerves, buccal lobe, visceral centers, and funnel nerves (Young, 1965a, 1967, 1971). A recent study provides evidence of marked somatotopy at the level of the basal lobe, where a defined topographical transform from the optic lobes has been identified in squids (Chung et al., 2020); this observation seems to parallel the case of insect and vertebrate brains, in which somatotopy is fairly ubiquitous.
iv. The presence of a blood-brain barrier, a unique property not found in other molluscs (Abbott and Pichon, 1987; for review see also Dunton et al., 2021).
v. Compound field potentials, similar to those recorded in vertebrate brains (e.g., Bullock and Budelmann, 1991; for review see Brown and Piscopo, 2013).
vi. An elevated efferent innervation of sensory receptors (e.g., the retina and equilibrium receptor organs, among others).
vii. The presence of peripheral first order afferent neurons (see: Young, 1971, 1991; Brown and Piscopo, 2013).
viii. A large variety of putative neurotransmitters and neuromodulators (review in Messenger, 1996; Ponte, 2012; Ponte and Fiorito, 2015).

During its evolution, the cephalopod brain achieved maximum aggregation and centralization of neural masses through fusion of the supra– and suboesophageal regions, which came to be enclosed in a cartilaginous cranium along with the expansion of two large optic lobes extending laterally from the supraoesophageal mass (directly behind the eyes). The most radical shift in the gross neural organization of the cephalopods resulted from a change in position and relative volume of the different areas of the nervous system that occurred with the addition or loss of ganglia. The accretion of fused ganglia ultimately yielded a central brain subdivided into a variable number of lobes (depending on species), ranging from 12 in the *Nautilus* to 24 in octopods (excluding the optic lobes). Notably, the central nervous system varies markedly across different cephalopod genera, with grades of neural complexity that parallel the density and complexity of sensory inputs received and the diversity of behaviors controlled and exhibited (Young, 1977a; Maddock and Young, 1987; Budelmann, 1995).

The greatest degree of nervous system centralization among cephalopods is found in the Octopodiformes, and is achieved by the shortening of the pathways connecting the superior buccal and brachial lobes (Nixon and Young, 2003). At the opposite end of the spectrum is the central nervous system of *Nautilus,* with three broad "bands" joining laterally (one dorsal and two ventral to the esophagus; Owen, 1832; Young, 1965b).

Overall, the octopod brain is more centralized than the decapod brain, in which brachial and pedal lobes are fused and the superior buccal lobe is united with the inferior frontal lobes. In addition, the brachial and pedal lobes of octopods, as well as their inferior frontal lobe system, are larger, reflecting the sophisticated use of their arms and highly elaborated chemotactile sensory processing and learning. Decapods, in contrast, have larger basal lobes and a simpler inferior frontal lobe system.

As enumerated by Ponte et al. (2021), different cephalopod brains manifest as taxon-specific 'cerebrotypes' akin to the specific types of brain architectures observed in the vertebrates. The significant quantitative differences between the brains of different cephalopod species reflect variations in habitat (in addition to other physical/environmental conditions). In the great majority of cases, the clusters of identified cerebrotypes correlate with similar ecological and/or behavioral constellations across different cephalopod species (Ponte et al., 2021). Within a total of 52 cephalopod species for which the set of data resulted complete, Ponte and coworkers recognized 10 distinctive groups of species, revealing both differences and close analogies. The overall topology of the relationships among species supports Young's perspective (Young, 1977a) and the working hypothesis that analyses combining relative brain size and life strategies can provide a robust basis for assumptions regarding the selective pressures and adaptations that drove cephalopod evolution. The analysis of cephalopod cerebrotypes (Ponte et al., 2021) highlights a large variation in the relative proportions of brain lobes within the decapods, as well as notable differences in the vertical lobe system when compared to that of the octopods. In fact, *O. vulgaris* presents a vertical lobe made up of *five* folded lobules that produce an overall volume reduction of the structure increasing the surface area and the corresponding number of cells in the lobe. This organization also results in reduction of the neuropilar space, minimization of the length of connections, increase in overall connectivity and computational abilities, akin to that observed in the higher vertebrates (Young, 1963, 1991, 1995; Shigeno et al., 2018). The opposite is true for cuttlefish and squid, where there is no observable folding of the surface of the vertical lobe, the estimated number of cells is much lower, and a correspondingly larger neuropil is found (Ponte et al., 2021).

A close relationship between cerebrotypes and lifestyles in cephalopods has thus been observed, supporting the idea that taxa evolved different sensory and cognitive strategies to cope with the differential demands of life in the ocean (Packard, 1972; Amodio et al., 2019; Ponte et al., 2021; Schnell et al., 2021c). Such complexity and diversity evoke comparisons to similar adaptations found among vertebrates. Taken together, these data support the idea that the appearance of cephalopod cerebrotypes reflect: (i) phylogenetic relationships (e.g., closely related species are likely to have a similar brain composition); (ii) similar developmental trajectories across different species (i.e., paralarvae vs. miniature adults at hatching) and constraints that influence brain organization and function; (iii) ecologically driven behavior which has led to the occupation of similar niches by species that possess similar brain architectures and faculties.

Though certainly noteworthy, the diversity of cerebrotypes is not the sole indicator of cephalopod brain complexity. In a recent review, Shigeno et al.and colleagues sought to establish structural and functional analogies to aspects of the vertebrate brain in the cephalopod nervous system. They undertook an analysis of the sensory, motor, and neurosecretory centers observed in cephalopod brains and attempted to identify «similarities to the

cerebral cortex, thalamus, basal ganglia, midbrain, cerebellum, hypothalamus, brain stem, and spinal cord of vertebrates» (Shigeno et al., 2018; see also Table 1 therein). The cephalopod cerebral cord can be considered analogous to the vertebrate forebrain and midbrain, while the pedal and palliovisceral cords are comparable to the vertebrate spinal cord and hindbrain. Evidence for other functional analogs of vertebrate brain features is steadily accumulating. Some examples are discussed below.

First, the existence of a functional analog of the hypothalamus is supported by the presence of neurosecretory cells in different lobes of the cephalopod brain. In vertebrates, the hypothalamus contains a population of neurosecretory cells, among other cell types (Butler and Hodos, 2005). Their evolutionary origins are believed to trace back to a common bilaterian ancestor, perhaps even a pre-bilaterian animal such as a cnidarian (Tessmar-Raible, 2007; Tessmar-Raible et al., 2007). In cephalopods, neurosecretory cells are found mainly in the buccal and sub-pedunculate lobes, as well as in some regions of the dorsal basal lobes, structures which all belong to the supra-esophageal mass (Young, 1970). Other areas reveal potential neurosecretory activity (i.e., sub-buccal and sub-pedunculate, optic gland, the neurovenous tissue of the vena cava; Bogoraze and Cazal, 1946; Barber, 1967; Young, 1970). Some of these regions are candidates for pituitary-hypothalamus analogs in the cephalopod brain, also presenting a subset of neurons containing molecules that are abundant in the hypothalamus, including GnRH and the vasopressin orthologs octopressin and cephalotocin (for review see Shigeno et al., 2018).

Second, the presence of higher sensory centers analogous to the thalamus has recently been proposed (Shigeno et al., 2018). The thalamus is the sensory relay center through which the majority of sensory inputs (excluding olfactory afferents) are directed to the mammalian cerebral cortex or non-mammalian vertebrate pallium (Swanson, 2007). The thalamus acts as a gatekeeper to the cortex and plays a key role in the perception of pain and, of particular note here, the generation of conscious states (Schiff, 2008; Rajneesh and Bolash, 2018; Redinbaugh et al., 2020). The cephalopod dorsal basal- and sub-vertical lobes are considered as candidate analogs of the vertebrate thalamus, as both receive numerous input fibers from the entire body *via* direct and indirect pathways from the sub-esophageal mass, thus acting together as a relay center for the outermost (i.e., cortically disposed) frontal and vertical lobes (Young, 1971). Although at least 10 major tracts originating from and/or terminating at the two structures have been identified in *O. vulgaris* (Young, 1971), to the best of our knowledge no estimation of the number of neural fibers comprising these tracts is available (but see Plän, 1987). Based on its dense connectivity, the dorsal basal lobe has also been proposed as a higher/intermediate motor center.

Furthermore, as discussed by Shigeno et al. (2018), the inferior frontal lobe appears to be another interesting candidate for sensory-motor integration, as a processing center for chemotactile information originating from lower centers (i.e., suckers on the arms), just as the olfactory cortex processes information from the olfactory receptors in vertebrates. Similar to its putative vertebrate counterpart, the inferior frontal lobe is part of the distributed neural matrix involved in learning and memory recall (the so-called chemo-tactile memory system; Young, 1991, 1995). Homologous structures have been identified in the brains of other cephalopods, and future efforts to uncover differences (if any) in the connectivity of the central neural structures of decapods and octopus may provide further insight.

Third, analogs of the vertebrate basal ganglia may be found in the higher motor centers of coleoid cephalopods (Young, 1971, 1977b). In particular, the anterior basal lobes (e.g., supra-esophageal mass) seem to exhibit analogous organization and function. Analysis of their neural connectivity, together with lesion experiments, support such an analogy (Chichery and Chichery, 1987; Gleadall, 1990). Considering their relative location, principal/major connectivity, functional organization (e.g., similarly hierarchical, progressing from motor pattern learning to central pattern controllers, initiators, generators, and motor neuron pools), these lobes are surmised to be plausible functional analogs of their vertebrate counterparts (Shigeno et al., 2018). Such higher motor centers receive sensory inputs and produce responses which, passing through the 'lower' parts of the central nervous system, are able to regulate posture, orientation, breathing, autonomic control of the viscera, and also habit formation (Shigeno et al., 2018). Analogs of vertebrate basal ganglia and their connections have been identified in different bilaterians (e.g., insects, annelids, and other protostomes) and seem to correspond to the basal lobe systems of cephalopods. However, functional analogies of such structures across taxa are not certain and each motor center has evolved specializations to meet the demands of a specific animal lineage, resulting in different body plans, locomotor systems and lifestyles across these taxa (Shigeno et al., 2018).

Though cortical structures are indeed fundamental for producing conscious states in mammals, subcortical areas are also essential, as they afford the integration of incoming signals into unified percepts and, ultimately, complex motor actions (Afrasiabi et al., 2021).

Given our limited knowledge of the function of the cephalopod basal lobes, as well as insufficiently supported claims regarding the existence of central pattern generators in these animals, we can only encourage further research in this direction.

Fourth and last, we focus on the associative (or auxiliary) centers of cephalopod brain as possible analogs of the vertebrate pallium or mammalian cerebral cortex.

In some cephalopods (e.g., *S. officinalis* and *O. vulgaris*) experimental evidence for sleep (Brown et al., 2006; Meisel et al., 2011; Frank et al., 2012; Iglesias et al., 2019; Medeiros et al., 2021), decision-making (see for example: Maldonado, 1963b, 1965; Carls-Diamante, 2017; Marini et al., 2017; Mather and Dickel, 2017), discrimination learning (for review see: Sanders, 1975; Boal, 1996; Marini et al., 2017), and structural and behavioral lateralization (Jozet-Alves et al., 2012a,b; Schnell et al., 2016a, 2018; Frasnelli et al., 2019) suggests a highly elaborated suite of cognitive faculties. It is not at all inconceivable that such a rich cognitive repertoire would require a neural substrate akin to the mammalian cortex (Edelman and Seth, 2009; Roth, 2015). As reviewed by Shigeno et al. (2018), an extensive series of experiments based on the ablation of different brain areas, followed by behavioral assays, revealed that the frontal and

vertical lobe systems (mainly in octopus, but also in cuttlefish) are involved in tactile and visual memory processing. As mentioned earlier, these structures contain large populations of uniquely distributed small interneurons (amacrine cells), parallel-running fibers, and reverberating circuitry across different lobes (Young, 1971, 1979, 1991, 1995). Notably, these are also areas in which synaptic, NMDA-independent long-term potentiation (LTP) has been discovered and characterized (Hochner et al., 2003; Shomrat et al., 2008, 2011; Turchetti-Maia et al., 2017). In addition, these lobe-systems appear to be characterized by heterogeneity of neurochemical identity (Ponte and Fiorito, 2015; Shigeno and Ragsdale, 2015). Further experiments are needed to assess cellular diversity and layered organization (especially of amacrine cells) in the frontal and vertical lobes, though preliminary data, including single cell sequencing, transcriptomes and molecular fingerprints related to learning outcomes (Zarrella, 2011; Zarrella et al., 2015; Manzo, 2021) strongly support this working hypothesis.

NEUROPHYSIOLOGICAL DYNAMICS AND THE FUNCTIONAL SIGNATURES OF CONSCIOUS STATES

Electrical activity in cephalopod brain has been assessed through various means and in different contexts (Bullock, 1984; Bullock and Budelmann, 1991; Brown et al., 2006), most recently in the characterization of neural activity (Butler-Struben et al., 2018). In a series of experiments, high-gain bipolar recordings obtained in the dorsal side of cephalopod brain (e.g., vertical lobe) and neighboring structures including the optic lobes (*via* electrodes inserted below the cartilaginous capsule) were able to capture organized electrical activity (Bullock, 1984; Brown et al., 2006). Recordings of brain signals using this methodology, and a similar one adopted by Butler-Struben et al. (2018), revealed periods of relative inactivity, as well as both spontaneous and evoked potentials. Interestingly, spontaneous activity in the areas within the vertical lobe is represented by single spikes and spike trains which are more frequent during rest, indicating a "body off/brain on" type of activation (Brown et al., 2006). Spike trains can last for tens of seconds, with frequencies ranging from about 10 to 40 Hz. The vertical lobe system is involved in learning and memory processing, displaying a vertebrate-like (albeit NMDA-independent) LTP plasticity (for review see Shomrat et al., 2015). Evidence of LTP-like plasticity has also been assessed *in vivo*. The signals recorded in this area in resting animals are believed to be related to memory consolidation, as in the vertebrate case (Shomrat et al., 2008, 2011). Compound potentials can also be evoked robustly in the optic lobes in response to brief flashes of light (Bullock, 1984). Their presence may be related to more basal levels of functional responsiveness of the nervous system to external stimuli.

Notably, the only experimental work involving the exposure of cephalopods to electroconvulsive shock (ECS) was a study by Maldonado (1968, 1969) in *O. vulgaris*. A two-second duration of ECS produced a general paroxysm of muscle contraction, inking, and cessation of breathing, as well as a flattening of the body with a strong adhesion of the suckers to the bottom of the box where the animals were placed. Once the animals were returned to their home tanks, they were initially completely rigid and immobile; breathing resumed shortly thereafter. Normal body posture and locomotor activities were restored within 15 min, and octopuses resumed their normal predatory responses within 2 h following the experiments (Maldonado, 1968, 1969). Interestingly, these studies were employed to assess the impairment of ECS on memory recall, further confirming the existence of sophisticated higher brain function, including the highly conserved biological machinery underlying long term memory.

The foregoing has been interpreted as psychological evidence of compound field potentials in cephalopods that are markedly different than those recorded in other invertebrates. In fact, cephalopod EEGs bear a close resemblance to vertebrate field potential recordings.

As summarized by Amodio and Fiorito (2013), one of the possible constraints on social learning in *O. vulgaris* is the lack of cross-modal integration, i.e., the ability to integrate stimuli from two or more sensory channels (for review see: Borrelli and Fiorito, 2008; Marini et al., 2017; namely, visual- and chemotactile-sensory motor systems). This is especially evident in instances where the solution to a task requires integration of the two modalities, as in the case of certain types of problem solving (Fiorito et al., 1990; Anderson and Mather, 2007; Anderson et al., 2008; Amodio and Fiorito, 2013). However, integration of different sensory channels is clearly demonstrated in foraging activities (Mather, 1991; Mather and O'Dor, 1991) as well as during social recognition, where sight, touch, and olfaction may be part of a multimodal system of information transfer (Partan and Marler, 2005; for examples in octopus see: Tricarico et al., 2011, 2014). Thus, synchronous use of different modalities (i.e., multimodality, Rowe and Guilford, 1999) has the clear advantage of improving detection, recognition, discrimination, and memorization of signals by the receivers, as recently shown in cuttlefish and octopus (Scheel et al., 2016; Schnell et al., 2016b).

Notably, Billard et al. (2020a) demonstrated the ability of cuttlefish to discriminate between and integrate two sensory modalities. Young concluded that complete integration (e.g., transfer) between two (visual and tactile information) systems occurred only at the effector level. However, Allen et al. (1986) showed that a limited degree of cross-modality does exist and the two sensory-motor systems may effectively integrate within higher neural centers: a finding recently supported by behavioral evidence provided by Kawashima et al. (2021).

Neurophysiological investigations have confirmed the view that cuttlefish and octopus evolved neural networks and synaptic plasticity paralleling the classic cellular basis of learning in mammals, i.e., LTP (Hochner et al., 2003; Shomrat et al., 2008, 2011; Hochner and Shomrat, 2013; Turchetti-Maia et al., 2017). However, in terms of architecture and physiological connectivity, the neural substrates for learning and memory in cephalopods evolved in a manner radically different than that of mammalian system (Shigeno et al., 2015), though functional properties analogous to those of mammalian cortical structures still emerged (e.g., limbic lobe as suggested by Young, 1995;

Shigeno et al., 2018). These structures - together constituting the vertical lobe system - are characterized by large populations of small nerve cells (e.g., amacrine cells) acting as interneurons which create highly redundant connections working *via en passant* innervations. This feature confers the octopus brain with the ability to create large-capacity memory associations (for review see for example: Sanders, 1975; Young, 1991; Shomrat et al., 2011, 2015; Hochner and Shomrat, 2013; Ponte and Fiorito, 2015). The complexity of neural circuitry is complemented by the rich diversity of neural cell types (e.g., Ponte, 2012; Ponte and Fiorito, 2015; Shigeno and Ragsdale, 2015; Deryckere et al., 2021), with a broad and specific differentiation among areas largely dominated by acetylcholine, catecholamines (dopamine and noradrenaline), indolamines (histamine, 5-HT), octopamine, purines, amino acids, nitric oxide, substance P, somatostatin, FMRF-amide, and other peptides which orchestrate responses at the level of the central and peripheral nervous systems, sensory organs, and viscera of cephalopods (Messenger, 1996). As reviewed by Ponte and Fiorito (2015), only limited regional differences among different neuromodulators appear to exist, and definite boundaries and/or mixing of cellular types have not been identified yet. Moreover, the complex distribution of different cell types in cephalopod brains is far from being characterized in any detail (Ponte, 2012; Ponte and Fiorito, 2015).

The various forms of learning and memory exhibited by cephalopods, the richness and flexibility of their behavioral repertoire (Borrelli and Fiorito, 2008; Marini et al., 2017; Hanlon and Messenger, 2018), and the unique adaptations and operating principles of the neural circuitry underlying their behavioral responses (Hochner et al., 2006; Shomrat et al., 2008, 2011, 2015; Hochner, 2012; Turchetti-Maia et al., 2017; Shigeno et al., 2018) should plausibly be considered markers for the presence of primary consciousness as proposed by Mather (2008).

CONCLUDING REMARKS

The sophisticated behavioral repertoire and cognitive abilities of cephalopod molluscs (Godfrey-Smith and Lawrence, 2012; Amodio and Fiorito, 2013; Tricarico et al., 2014; Scheel et al., 2016) strongly suggest the presence of conscious states in these animals, as further enunciated during the recent well-articulated debate attending the notion of cephalopod 'mind' (see Mather, 2019)[1] which included contributions from philosophers and professionals from artistic and cultural domains. While discussion surrounding the attribution of consciousness in cephalopods is still ongoing, the growing body of evidence that, at the very least, it would be prudent to apply the precautionary principle, as implied by the thrust of the present work.

The extraordinary behavioral and cognitive features that cephalopods possess (Godfrey-Smith, 2013, 2016; Marini et al., 2017; Hanlon and Messenger, 2018; Gutnick et al., 2021) have long attracted the public's imagination (e.g., Nakajima, 2018; Nakajima et al., 2018; Holden-Dye et al., 2019). When we consider the neural hallmarks of consciousness (Edelman et al., 2005; Seth et al., 2005; Edelman and Seth, 2009; Edelman, 2011), we must take into account morphological and functional analogies (Young, 1991, 1995; Edelman and Seth, 2009; Albertin et al., 2015; Shigeno et al., 2015; see also Shigeno et al., 2018) which reinforce the argument that nature often achieves the same goals across phylogeny in a number of different ways, some of which may accord with current anatomical and physiological views of how the neural systems underlying complex behavior actually works (see, e.g., Rankin, 2004). It may be useful to recall the argument for biological convergence that was made by Edelman et al. (Edelman et al., 2005; Seth et al., 2005; see also: Edelman and Seth, 2009; Boly et al., 2013) as part of a synthetic approach to the study of animal consciousness. Those Authors argued that, in the assessment of possible conscious states in non-human species, a comparative examination of neuroanatomical, neurophysiological, and behavioral properties and correlates using the human case as a kind of reference standard could provide a way forward. Entertaining the possibility that the phylogeny of consciousness might include some invertebrate lines, they further posited that, even in the absence of neuroanatomy that is structurally *homologous* to that of vertebrates, it is possible that some invertebrates evolved aspects of brain architecture that are functionally *analogous* to neural structures and circuits critical to instantiating conscious states in vertebrates. The fact that invertebrate nervous system do not possess anything that *looks* like cortex, hippocampus, or thalamus does not mean - as we have seen above - that cephalopods are not equipped with specialized structures and circuitry that support *similar functions*, i.e., working and episodic-like memory, storage, and retrieval (or recall) akin to those faculties supported by cortex and hippocampus, as well as recursive - or reentrant - relays that link perception and memory in a manner similar to that afforded by the vertebrate thalamus (see Edelman, 1987, 1989).

In addition to indirect circumstantial evidence of consciousness in cephalopods provided by the outstanding flexibility of their behavioral repertoire, and their relatively complex and specialized neural structures instantiating circuitry resembling that found in vertebrates, the likelihood of consciousness in these invertebrates is supported by more reliable objective, basic neural correlates, such as EEG-like signatures and evoked compound potentials. In acknowledging these facts, the Cambridge Declaration on Consciousness[2] recognized cephalopods as animals whose neurobiological structures are complex enough to support conscious states. Furthermore, Directive 2010/63/EU has included cephalopods as the sole species among invertebrates listed among the animals whose welfare should be protected for scientific research (Smith et al., 2013; Fiorito et al., 2014, 2015).

Birch et al. suggest that animal consciousness could be conceptualized whitin a broader framework consisting of five dimensions that do not force species into hierarchical positions of higher versus lower levels of consciousness, but rather consider

[1] See also the article thread available at https://www.wellbeingintlstudiesrepository.org/animsent/vol4/iss26/1/

[2] https://web.archive.org/web/20131109230457/http://fcmconference.org/img/CambridgeDeclarationOnConsciousness.pdf

species within their own space: a paradigm that helps us better understand their unique abilities and cognitive profiles without imposing meaningless comparisons (Birch et al., 2020). After all, we cannot expect an octopus to experience the world in the same way that we do. Our sensorimotor systems and the environments we inhabit are radically different and our evolutionary histories are quite divergent. As reviewed above and summarized in **Table 1**, the five dimensions (Birch et al., 2020) and the hallmarks of consciousness as possible counterparts (Edelman et al., 2005; Seth et al., 2005; Edelman and Seth, 2009) together incorporate perceptual and evaluative richness, integration at both a point in time and over time, and self-awareness (though the distinction between the latter as a higher-order form of consciousness and primary, or sensory, consciousness should be noted).

P-richness refers to the different level of detail with which animals consciously perceive aspects of their environment. Of course, as noted above, this varies according to the sensory systems with which each species is endowed (e.g., chemical-tactile, visual, and auditory). Cephalopods appear to possess a large p-richness in chemo-tactile and visual discrimination (review in: Marini et al., 2017; Mather, 2021b) and are able to retain episodic-like memories (e.g., Pronk et al., 2010; Jozet-Alves et al., 2013).

E-richness refers to the differential affective experience of animals in relation to particular stimuli, and thus to the ability to detect negative or positive valence (in cephalopods see for example: Maldonado, 1963b, 1965; Darmaillacq et al., 2004), which are of course determined by different species- and age-specific physiological needs and motivations. Cephalopods are also likely to have good e-richness, as there is accumulating evidence suggesting the presence of nociception and pain in these animals (Crook et al., 2011, 2013; Alupay et al., 2014; Oshima et al., 2016; Crook, 2021). Unity and Temporality are closely related to how animals subjectively perceive their environments in relation to time and whether they are able to remember, retain, and retrieve information over time (see discussion above). Selfhood refers to an animal's ability to distinguish itself from the outside world and from others (e.g., mirror test).

As summarized in **Table 1**, sensory-motor communication in the brain of multiple sensorial inputs (p-richness) is the backbone upon which the unity of time-coding, self-awareness, arousal, and motivation are instantiated. In cephalopods the mechanisms of attention and decision making, modulated by D1 or D2 neuronal types in the mammalian striatum, are still unclear. However, the analogies with the mammalian basal ganglia mentioned above, as well as the existence of an intricate dopaminergic (and octopaminergic) network with spatial distribution in specific brain areas (Ponte, 2012; Ponte and Fiorito, 2015) are also indicators of e-richness in cephalopods. Further investigations of possible cephalopod analogs of the cortico-basal ganglia pathways and basal ganglia-thalamic neural pathways will be required to experimentally advance our overview. Gene editing, as recently promoted in cephalopods (Crawford et al., 2020; Steele, 2020) may also help over this challenging avenue.

In mammals, the combination of connectivity-based optogenetic tagging and psychophysical approaches has been pivotal for revealing how interactions between the thalamus and cortex control the sensory and limbic processing that underlies higher cognitive functions (Halassa et al., 2014). Optogenetic studies in mice have allowed the identification of intricate neural networks, possibly contributing to mechanisms of consciousness, including pathways originating from the striatum that inhibit the thalamic reticular nucleus and participate in the regulation of arousal, decision making and states of consciousness (Halassa et al., 2014; Halassa and Kastner, 2017; Schmitt et al., 2017). Optogenetic approaches are in their early infancy in cephalopods, but their potential has been recently exploited with success (Reiter et al., 2018; Reiter and Laurent, 2020). We are convinced that further studies will benefit from an integration of approaches.

Though based on incomplete behavioral, morphological, and physiological findings (thus, considering the precautionary principle; EFSA Panel, 2005), cephalopods have been included in Directive 2010/63/EU as the only invertebrates among the so-called laboratory animals to be protected in scientific research. Originally adopted within the context of environmental law, the 'precautionary principle' is based on the idea that in cases of threat of actual or potential irreversible damage to the environment, the lack of complete scientific evidences should not be employed as a reason for postponing measures to be taken in order to avoid or minimize the risks (Cameron and Abouchar, 1991; Pinto-Bazurco, 2020). In respect to animals and their welfare, the same principle has been adopted (EFSA Panel, 2005; Andrews, 2011) even employing sentience (Birch, 2017) and consciousness (Bradshaw, 1998; Dawkins, 2017) as justifications. In the words of Bradshaw «Applying this principle [i.e., precautionary] to the issue of animal consciousness, the following rule is formulated: assume animals do have consciousness in case they do; if they do not it does not matter» (1998, p. 108).

It is now evident that adopting a multidimensional approach has completely changed our perspective on animal consciousness and has made us realize that we may have been asking the wrong question, namely "is this species more conscious than that one?," when the more relevant question should be: "how is the individual experience of this species different from that one?"

The five dimensions enumerated by Birch et al. (2020) are, to some extent, included in the definition of what could be considered the 'anteroom' of consciousness in animals, namely sentience. According to Broom (2014), a sentient being has at least one of the following abilities: (i) evaluation of the actions of others in relation to itself (e.g., the capacity to form relationships); (ii) the capacity to remember some of one's own actions and their consequences (e.g., cognitive ability); (iii) the ability to assess risks and benefits (e.g., decision-making); (iv) possession of some degree of awareness (e.g., consciousness); (v) the ability to experiencing negative or positive affective states (e.g., the influence of others' states).

Based on the available evidence - reviewed in the present work - we believe that cephalopods are sentient animals in terms of all five capacities summarized above. It will be both intriguing and enlightening to dissect sentience from the cephalopod perspective, based on knowledge accumulated over several decades, as well as on recently gathered evidence and arguments that support the invocation of EFSA guidelines for

the inclusion of this taxon in the list of species regulated by the Directive 2010/63/EU (EFSA Panel, 2005; European Parliament Council of the European Union, 2010). But this is a pursuit best reserved for a different time and venue.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

GP, CC, and GF conceived this work. DE contributed to writing together with EP and PI. All authors discussed the content, text and contributed to writing, commented on the manuscript at all stages, and read and approved the submitted manuscript.

FUNDING

This work has been supported by the Stazione Zoologica Anton Dohrn and Association for Cephalopod Research CephRes.

ACKNOWLEDGMENTS

A portion of this manuscript originated from a presentation by one of us (GF) and the ensuing discussion at the 12th International Symposium on the Science of Behavior (ISSB, Brooklyn College, July 20–22, 2015). We are indebted to the Organizers (University of Guadalajara - Mexico and Brooklyn College – NY, USA) and colleagues who attended the symposium for their contribution to a warm reception and fascinating discussion. We would also like to acknowledge Elena Tricarico for suggestions and the contribution by Dr. Fabio De Sio to the discussion of theory of mind.

REFERENCES

Abbott, N. J., and Pichon, Y. (1987). The glial blood-brain barrier of crustacea and cephalopods: a review. *J. Physiol. Paris* 82, 304–313.

Adamo, S. A., Ehgoetz, K., Sangster, C., and Whitehorne, I. (2006). Signaling to the enemy? Body Pattern Expression and Its Response to External Cues During Hunting in the Cuttlefish Sepia officinalis (Cephalopoda). *Biol. Bull.* 210, 192–200. doi: 10.2307/4134557

Afrasiabi, M., Redinbaugh, M. J., Phillips, J. M., Kambi, N. A., Mohanta, S., Raz, A., et al. (2021). Consciousness depends on integration between parietal cortex, striatum, and thalamus. *Cell Syst. 12,* 363–373.e311. doi: 10.1016/j.cels.2021.02.003

Agin, V., Chichery, R., Dickel, L., and Chichery, M. P. (2006). The "prawn-in-the-tube" procedure in the cuttlefish: habituation or passive avoidance learning? *Learn. Memory* 13, 97–101. doi: 10.1101/lm.90106

Albertin, C. B., Simakov, O., Mitros, T., Wang, Z. Y., Pungor, J. R., Edsinger-Gonzales, E., et al. (2015). The octopus genome and the evolution of cephalopod neural and morphological novelties. *Nature* 524, 220–224. doi: 10.1038/nature14668

Allen, A., Michels, J., and Young, J. Z. (1986). Possible interactions between visual and tactile memories in octopus. *Mar. Behav. Physiol.* 12, 81–97. doi: 10.1080/10236248609378636

Alupay, J. S., Hadjisolomou, S. P., and Crook, R. J. (2014). Arm injury produces long-term behavioral and neural hypersensitivity in octopus. *Neurosci. Lett.* 558, 137–142. doi: 10.1016/j.neulet.2013.11.002

Amodio, P., Andrews, P., Salemme, M., Ponte, G., and Fiorito, G. (2014). The use of artificial crabs for testing predatory behavior and health in the octopus. *Altex Alternat. Anim. Exp.* 31, 494–499. doi: 10.14573/altex.1401282s

Amodio, P., Boeckle, M., Schnell, A. K., Ostojíc, L., Fiorito, G., and Clayton, N. S. (2019). Grow smart and die young: why did cephalopods evolve intelligence? *Trends Ecol. Evol.* 34, 45–56. doi: 10.1016/j.tree.2018.10.010

Amodio, P., and Fiorito, G. (2013). "Observational and other types of learning in octopus," in *Invertebrate Learning and Memory*, eds R. Menzel and P. Benjamin (London: Academic Press), 293–302.

Anderson, R. C., and Mather, J. A. (2007). The packaging problem: bivalve prey selection and prey entry techniques of the octopus *Enteroctopus dofleini*. *J. Comp. Psychol* 121, 300–305. doi: 10.1037/0735-7036.121.3.300

Anderson, R. C., and Mather, J. A. (2010). It's all in the cues: octopuses (*Enteroctopus dofleini*) learn to open jars. *Ferrantia* 59, 8–13.

Anderson, R. C., Mather, J. A., Monette, M. Q., and Zimsen, S. R. (2010). Octopuses (*Enteroctopus dofleini*) recognize individual humans. *J. Appl. Anim. Welfare Sci.* 13, 261–272. doi: 10.1080/10888705.2010.483892

Anderson, R. C., Sinn, D. L., and Mather, J. A. (2008). Drilling localization on bivalve prey by *Octopus rubescens* Bery, 1953 (Cephalopoda: Octopodidae). *Veliger* 50, 326–328.

Andrews, K. (2020). *How to Study Animal Minds (Elements in the Philosophy of Biology)*. Cambridge: Cambridge University Press.

Andrews, P. L. R. (2011). Laboratory invertebrates: only spineless, or spineless and painless? *Introduction. ILAR J.* 52, 121–125. doi: 10.1093/ilar.52.2.121

Apperly, I. (2011). Mindreaders. The Cognitive basis of "Theory of Mind". Hove: Psychology Press.

Aristotle (1910). *Historia Animalium, english translation by D'Arcy Wenthworth Thompson*. Oxford: Clarendon Press.

Avarguès-Weber, A., D'amaro, D., Metzler, M., Garcia, J., and Dyer, A. (2017). Recognition of human face images by the free flying wasp *Vespula vulgaris*. *Anim. Behav. Cogn.* 4, 314–323. doi: 10.26451/abc.04.03.09.2017

Baars, B. J. (1994). "A global workspace theory of conscious experience," in *Consciousness in Philosophy and Cognitive Neuroscience*, eds A. Revonsuo and M. Kamppinen (New York, NY: Psychology Press), 161–184.

Baars, B. J. (2002). The conscious access hypothesis: origins and recent evidence. *Trends Cogn. Sci.* 6, 47–52. doi: 10.1016/S1364-6613(00)01819-2

Barber, V. C. (1967). A neurosecretory tissue in octopus. *Nature* 213, 1042–1043. doi: 10.1038/2131042a0

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., and Plumb, I. (2001). The "Reading the Mind in the Eyes" test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *J. Child Psychol. Psychiatry* 42, 241–251. doi: 10.1111/1469-7610.00715

Bayne, T. (2010). *The Unity of Consciousness.* New York, NY: Oxford University Press.

Billard, P., Clayton, N. S., and Jozet-Alves, C. (2020a). Cuttlefish retrieve whether they smelt or saw a previously encountered item. *Sci. Rep.* 10, 5413. doi: 10.1038/s41598-020-62335-x

Billard, P., Schnell, A. K., Clayton, N. S., and Jozet-Alves, C. (2020b). Cuttlefish show flexible and future-dependent foraging cognition. *Biol. Lett.* 16, 20190743. doi: 10.1098/rsbl.2019.0743

Birch, J. (2017). Animal sentience and the precautionary principle. *Anim. Sentience* 2, 1. doi: 10.51291/2377-7478.1200

Birch, J., Burn, C., Schnell, A., Browning, H., and Crump, A. (2021). *Review of the Evidence of Sentience in Cephalopod Molluscs and Decapod Crustaceans*. Available online at: https://philpapers.org/rec/BIRROT-5

Birch, J., Schnell, A. K., and Clayton, N. S. (2020). Dimensions of animal consciousness. *Trends Cogn. Sci.* 24, 789–801. doi: 10.1016/j.tics.2020.07.007

Boal, J. G. (1996). A review of simultaneous visual discrimination as a method of training octopuses. *Biol. Rev.* 71, 157–190. doi: 10.1111/j.1469-185X.1996.tb00746.x

Boal, J. G. (2006). Social recognition: a top down view of cephalopod behaviour. *Vie Milieu* 56, 69–79.

Bogoraze, D., and Cazal, P. (1946). Remarques sur le système stomatogastrique du Poulpe (*Octopus vulgaris* Lamarck): le complexe retro-buccal. *Arch. zoologie expérimentale générale* 84, 115–131.

Boivin, X., Nowak, R., Despres, G., Tournadre, H., and Le Neindre, P. (1997). Discrimination between shepherds by lambs reared under artificial conditions. *J. Anim. Sci.* 75, 2892–2898. doi: 10.2527/1997.75112892x

Boly, M., Seth, A., Wilke, M., Ingmundson, P., Baars, B., Laureys, S., et al. (2013). Consciousness in humans and non-human animals: recent advances and future directions. *Front. Psychol.* 4, 625. doi: 10.3389/fpsyg.2013.00625

Borrelli, L. (2007). *Testing the contribution of relative brain size and learning capabilities on the evolution of Octopus vulgaris and other cephalopods* (PhD Thesis). Stazione Zoologica Anton Dohrn, Italy; Open University, United Kingdom.

Borrelli, L., Chiandetti, C., and Fiorito, G. (2020). A standardized battery of tests to measure *Octopus vulgaris*' behavioural performance. *Invertebrate Neurosci.* 20, 4. doi: 10.1007/s10158-020-0237-7

Borrelli, L., and Fiorito, G. (2008). "Behavioral analysis of learning and memory in cephalopods," in *Learning and Memory: A Comprehensive Reference*, ed J. J. Byrne (Oxford: Academic Press), 605–627.

Borrelli, L., Gherardi, F., and Fiorito, G. (2006). *A Catalogue of Body Patterning in Cephalopoda. Napoli: Stazione Zoologica A. Dohrn; Firenze University Press.*

Boycott, B. B. (1954). Learning in *Octopus vulgaris* and other cephalopods. *Pubbl. Staz. Zool. Napoli* 25, 67–93.

Bradshaw, R. H. (1998). Consciousness in non-human animals: adopting the precautionary principle. *J. Consciousness Stud.* 5, 108–114.

Broom, D. M. (2014). *Sentience and Animal Welfare*. Oxfordshire: CAB International.

Brown, E. R., and Piscopo, S. (2013). Synaptic plasticity in cephalopods; more than just learning and memory? *Invertebrate Neurosci.* 13, 35–44. doi: 10.1007/s10158-013-0150-4

Brown, E. R., Piscopo, S., De Stefano, R., and Giuditta, A. (2006). Brain and behavioural evidence for rest-activity cycles in *Octopus vulgaris*. *Behav. Brain Res.* 172, 355–359. doi: 10.1016/j.bbr.2006.05.009

Bublitz, A., Weinhold, S. R., Strobel, S., Dehnhardt, G., and Hanke, F. D. (2017). Reconsideration of serial visual reversal learning in octopus (*Octopus vulgaris*) from a methodological perspective. *Front. Physiol.* 8, 54. doi: 10.3389/fphys.2017.00054

Budelmann, B. U. (1995). "The cephalopod nervous system: what evolution has made of the molluscan design," in *The Nervous Systems of Invertebrates: An Evolutionary and Comparative Approach*, eds O. Breidbach and W. Kutsch (Basel: Birkhäuser Verlag), 115–138.

Bugnyar, T., Stöwe, M., and Heinrich, B. (2004). Ravens, Corvus corax, follow gaze direction of humans around obstacles. *Proc. R. Soc. London B Biol. Sci.* 271, 1331–1336. doi: 10.1098/rspb.2004.2738

Bullock, T. H. (1984). Ongoing compound field potentials from octopus brain are labile and vertebrate-like. *Electroencephalogr. Clin. Neurophysiol.* 57, 473–483. doi: 10.1016/0013-4694(84)90077-4

Bullock, T. H., and Budelmann, B. U. (1991). Sensory evoked potentials in unanesthetized unrestrained cuttlefish: a new preparation for brain physiology in cephalopods. *J. Comp. Physiol. A* 168, 141–150. doi: 10.1007/BF00217112

Butler, A. B., and Hodos, W. (2005). *Comparative Vertebrate Neuroanatomy: Evolution and Adaptation*. Hoboken, NJ: John Wiley & Sons.

Butler-Struben, H. M., Brophy, S. M., Johnson, N. A., and Crook, R. J. (2018). *In vivo* recording of neural and behavioral correlates of anesthesia induction, reversal, and euthanasia in cephalopod molluscs. *Front. Physiol.* 9, 109. doi: 10.3389/fphys.2018.00109

Buytendijk, F. J. J. (1933). Das Verhalten von Octopus nach Teilweiser Zerstöring des "gehirns". *Arch. Néerl. Physiol.* 18, 24–70.

Caillois, R. (1973). *La pieuvre. Essai sur la logique de l'imaginaire*. Paris: La Table ronde.

Caldwell, R. L., Ross, R., Rodaniche, A., and Huffard, C. L. (2015). Behavior and body patterns of the larger pacific striped octopus. *PLoS ONE* 10, e0134152. doi: 10.1371/journal.pone.0134152

Calvé, M. R. (2005). *Individual Differences in the Common Cuttlefish, Sepia officinalis* (Master of Science). Department of Biology, Dalhousie University, Halifax, Canada.

Cameron, J., and Abouchar, J. (1991). The precautionary principle: a fundamental principle of law and policy for the protection of the global environment. *Boston College Int. Comp. Law Rev.* 14, 1–27.

Cannicci, S., Morino, L., and Vannini, M. (2002). Behavioural evidence for visual recognition of predators by the mangrove climbing crab *Sesarma leptosoma*. *Anim. Behav.* 63, 77–83. doi: 10.1006/anbe.2001.1882

Carls-Diamante, S. (2017). The octopus and the unity of consciousness. *Biol. Philos.* 32, 1269–1287. doi: 10.1007/s10539-017-9604-0

Carls-Diamante, S. (2021). *The Octopus: Implications for Cognitive Science*. Available online at: https://escholarship.org/uc/item/8sz9757q

Carter, J., Lyons, N. J., Cole, H. L., and Goldsmith, A. R. (2008). Subtle cues of predation risk: starlings respond to a predator's direction of eye-gaze. *Proc. R. Soc. London B Biol. Sci.* 275, 1709–1715. doi: 10.1098/rspb.2008.0095

Cartron, L., Darmaillacq, A. S., and Dickel, L. (2013). The "prawn-in-the-tube" procedure: what do cuttlefish learn and memorize? *Behav. Brain Res.* 240, 29–32. doi: 10.1016/j.bbr.2012.11.010

Chapko, M. K., Grossbeck, M. L., Hansen, R. L., Maher, T. D., Middleton, R. S., and Simpson, R. W. (1962). *Devilfish. A Practical Guide to the Dissection of Octopus.* Ontario Center: Wayne Senior High School.

Chase, R., and Wells, M. J. (1986). Chemotactic behavior in octopus. *J. Compar. Physiol. A Sensory Neural Behav. Physiol.* 158, 375–381. doi: 10.1007/BF00603621

Chiao, C.-C., and Hanlon, R. T. (2001a). Cuttlefish camouflage: visual perception of size, contrast and number of white squares on artificial checkerboard substrata initiates disruptive coloration. *J. Exp. Biol.* 204, 2119–2125. doi: 10.1242/jeb.204.12.2119

Chiao, C.-C., and Hanlon, R. T. (2001b). Cuttlefish cue visually on area-not shape or aspect ratio-of light objects in the substrate to produce disruptive body patterns for camouflage. *Biol. Bull.* 201, 269–270. doi: 10.2307/1543359

Chichery, M., and Chichery, R. (1987). The anterior basal lobe and control of prey-capture in the cuttlefish (*Sepia officinalis*). *Physiol. Behav.* 40, 329–336. doi: 10.1016/0031-9384(87)90055-2

Chung, W.-S., Kurniawan, N. D., and Marshall, N. J. (2020). Toward an MRI-based mesoscale connectome of the squid brain. *iScience* 23, 100816. doi: 10.1016/j.isci.2019.100816

Clayton, N. S., Bussey, T. J., and Dickinson, A. (2003). Can animals recall the past and plan for the future? *Nat. Rev. Neurosci.* 4, 685–691. doi: 10.1038/nrn1180

Clymer, J. C. (1973). *A Computer Simulation Model of Attack-Learning Behavior in the Octopus* (PhD), Chicago, MI: The University of Michigan.

Cole, P. D., and Adamo, S. A. (2005). Cuttlefish (*Sepia officinalis*: Cephalopoda) hunting behavior and associative learning. *Anim. Cogn.* 8, 27–30. doi: 10.1007/s10071-004-0228-9

Courage, K. H. (2013). *Octopus! The Most Mysterious Creature in the Sea*. New York, NY: Penguin.

Cousteau, J.-Y., and Diolé, P. (1973). *Octopus and Squid. The Soft Intelligence*. Garden City, NY: Doubleday and Co., Inc.

Craik, K. J. W. (1967). *The Nature of Explanation (1st Updated Edition of the Original, 1943)*. Cambridge: Cambridge University Press.

Crawford, K., Diaz Quiroz, J. F., Koenig, K. M., Ahuja, N., Albertin, C. B., and Rosenthal, J. J. C. (2020). Highly efficient knockout of a squid pigmentation gene. *Curr. Biol.* 30, 3484–3490.e3484. doi: 10.1016/j.cub.2020.06.099

Crook, R. J. (2021). Behavioral and neurophysiological evidence suggests affective pain experience in octopus. *iScience* 24, 102229. doi: 10.1016/j.isci.2021.102229

Crook, R. J., Hanlon, R. T., and Walters, E. T. (2013). Squid have nociceptors that display widespread Long-Term Sensitization and spontaneous activity after bodily injury. *J. Neurosci.* 33, 10021–10026. doi: 10.1523/JNEUROSCI.0646-13.2013

Crook, R. J., Lewis, T., Hanlon, R. T., and Walters, E. T. (2011). Peripheral injury induces long-term sensitization of defensive responses to visual and tactile stimuli in the squid *Loligo pealeii*, Lesueur 1821. *J. Exp. Biol.* 214, 3173–3185. doi: 10.1242/jeb.058131

Crook, R. J., and Walters, E. T. (2014). Neuroethology: self-recognition helps octopuses avoid entanglement. *Curr. Biol.* 24, R520–R521. doi: 10.1016/j.cub.2014.04.036

Crystal, J. D. (2010). Episodic-like memory in animals. *Behav. Brain Res.* 215, 235–243. doi: 10.1016/j.bbr.2010.03.005

Darmaillacq, A. S., Dickel, L., Chichery, M. P., Agin, V., and Chichery, R. (2004). Rapid taste aversion learning in adult cuttlefish, *Sepia officinalis*. *Anim. Behav.* 68, 1291–1298. doi: 10.1016/j.anbehav.2004.01.015

Darwin, C. (1870). *Journal of the Researches Into the Natural History and Geology of the Countries Visited During the Voyage of H. M. S. Beagle Round the World Under the Command of Capt. Fitz Roy, R. N*. London: John Murray.

Dawkins, M. S. (2017). Animal welfare with and without consciousness. *J. Zool.* 301, 1–10. doi: 10.1111/jzo.12434

De Sio, F. (2011). Leviathan and the soft animal: medical humanism and the invertebrate models for higher nervous functions, 1950s-90s. *Med. Hist.* 55, 369–374. doi: 10.1017/S0025727300005421

De Sio, F., Hanke, F. D., Warnke, K., Marazia, C., Galligioni, V., Fiorito, G., et al. (2020). E Pluribus octo - building consensus on standards of care and experimentation in cephalopod research; a historical outlook. *Front. Physiol.* 11, 645. doi: 10.3389/fphys.2020.00645

DeHaene, S., and Changeux, J.-P. (2004). "Neural mechanisms for access to consciousness," in *The Cognitive Neurosciences, 3rd Edn,* ed M. S. Gazzaniga (Cambridge, MA: Boston Review), 1145–1157.

Dehaene, S., and Changeux, J.-P. (2005). Ongoing spontaneous activity controls access to consciousness: a neuronal model for inattentional blindness. *PLoS Biol.* 3, e141. doi: 10.1371/journal.pbio.0030141

Deryckere, A., Styfhals, R., Elagoz, A. M., Maes, G. E., and Seuntjens, E. (2021). Identification of neural progenitor cells and their progeny reveals long distance migration in the developing octopus brain. *eLife* 10:e69161. doi: 10.7554/eLife.69161.sa2

Di Cristina, G., Andrews, P., Ponte, G., Galligioni, V., and Fiorito, G. (2015). The impact of Directive 2010/63/EU on cephalopod research. *Invertebrate Neurosci.* 15, 8. doi: 10.1007/s10158-015-0183-y

Dröscher, A. (2016). Pioneering studies on Cephalopod's Eye and Vision at the Stazione Zoologica Anton Dohrn (1883-1977). *Front. Physiol.* 7, 618. doi: 10.3389/fphys.2016.00618

Dunton, A. D., Göpel, T., Ho, D. H., and Burggren, W. (2021). Form and function of the vertebrate and invertebrate blood-brain barriers. *Int. J. Mol. Sci.* 22, 12111. doi: 10.3390/ijms222212111

Dyer, A. G., Neumeyer, C., and Chittka, L. (2005). Honeybee (*Apis mellifera*) vision can discriminate between and recognise images of human faces. *J. Exp. Biol.* 208, 4709–4714. doi: 10.1242/jeb.01929

Edelman, D. B. (2011). How octopuses see the world and other roads less traveled: necessity versus sufficiency and evolutionary convergence in the study of animal consciousness. *J. Shellfish Res.* 30, 1001.

Edelman, D. B., Baars, B. J., and Seth, A. K. (2005). Identifying hallmarks of consciousness in non-mammalian species. *Conscious. Cogn.* 14, 169–187. doi: 10.1016/j.concog.2004.09.001

Edelman, D. B., and Seth, A. K. (2009). Animal consciousness: a synthetic approach. *Trends Neurosci.* 32, 476–484. doi: 10.1016/j.tins.2009.05.008

Edelman, G., Gally, J., and Baars, B. (2011). Biology of consciousness. *Front. Psychol.* 2, 4. doi: 10.3389/fpsyg.2011.00004

Edelman, G. M. (1987). *Neural Darwinism: The Theory of Neuronal Group Selection*. New York, NY: Basic Books.

Edelman, G. M. (1989). *The Remembered Present: A Biological Theory of Consciousness*. New York, NY: Basic Books.

Edelman, G. M. (1993). Neural Darwinism: selection and reentrant signaling in higher brain function. *Neuron* 10, 115–125. doi: 10.1016/0896-6273(93)90304-A

Edelman, G. M. (2003). Naturalizing consciousness: a theoretical framework. *Proc. Nat. Acad. Sci. U.S.A.* 100, 5520–5524. doi: 10.1073/pnas.0931349100

EFSA Panel, o. A. H. a. W. (2005). Opinion of the Scientific Panel on Animal Health and Welfare (AHAW) on a request from the commission related to the "aspects of the biology and welfare of animals used for experimental and other scientific purposes". *EFSA J.* 292, 1–136. doi: 10.2903/j.efsa.2005.292

El Nagar, A., Osorio, D., Zylinski, S., and Sait, S. M. (2021). Visual perception and camouflage response to 3D backgrounds and cast shadows in the European cuttlefish, *Sepia officinalis*. *J. Exp. Biol.* 224, jeb238717. doi: 10.1242/jeb.238717

Emery, N. J., and Clayton, N. S. (2009). Comparative social cognition. *Annu. Rev. Psychol.* 60, 87–113. doi: 10.1146/annurev.psych.60.110707.163526

European Parliament and Council of the European Union (2010). Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the Protection of Animals Used for Scientific Purposes. Strasbourg: Concil of Europe. Available online at: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32010L0063 (last visited August 2020).

Feinberg, T. E., and Mallatt, J. (2020). Phenomenal consciousness and emergence: eliminating the explanatory gap. *Front. Psychol.* 11, 1041. doi: 10.3389/fpsyg.2020.01041

Feord, R. C., Sumner, M. E., Pusdekar, S., Kalra, L., Gonzalez-Bellido, P. T., and Wardill, T. J. (2020). Cuttlefish use stereopsis to strike at prey. *Sci. Adv.* 6, eaay6036. doi: 10.1126/sciadv.aay6036

Finn, J. K., Tregenza, T., and Norman, M. D. (2009). Defensive tool use in a coconut-carrying octopus. *Curr. Biol.* 19, R1069–R1070. doi: 10.1016/j.cub.2009.10.052

Fiorito, G. (1993). Social learning in invertebrates. *Response Sci.* 259, 1629. doi: 10.1126/science.259.5101.1629

Fiorito, G., Affuso, A., Anderson, D. B., Basil, J., Bonnaud, L., Botta, G., et al. (2014). Cephalopods in neuroscience: regulations, research and the 3Rs. *Invert. Neurosci.* 14, 13–36. doi: 10.1007/s10158-013-0165-x

Fiorito, G., Affuso, A., Basil, J., Cole, A., de Girolamo, P., D'Angelo, L., et al. (2015). Guidelines for the care and welfare of cephalopods in research -a consensus based on an initiative by CephRes, FELASA and the Boyd Group. *Lab. Anim.* 49, 1–90. doi: 10.1177/0023677215580006

Fiorito, G., and Chichery, R. (1995). Lesions of the vertical lobe impair visual discrimination learning by observation in *Octopus vulgaris*. *Neurosci. Lett.* 192, 117–120. doi: 10.1016/0304-3940(95)11631-6

Fiorito, G., and Scotto, P. (1992). Observational learning in *Octopus vulgaris*. *Science* 256, 545–547. doi: 10.1126/science.256.5056.545

Fiorito, G., von Planta, C., and Scotto, P. (1990). Problem solving ability of *Octopus vulgaris* lamarck (Mollusca, Cephalopoda). *Behav. Neural Biol.* 53, 217–230. doi: 10.1016/0163-1047(90)90441-8

Fisher, J. (1954). "Evolution and bird sociality," in *Evolution as a Process*, eds J. Huxley, A. C. Hardy, and E. B. Ford (London: Allen & Unwin), 71–83.

Forsythe, J. W., and Hanlon, R. T. (1997). Foraging and associated behavior by *Octopus cyanea* Gray, 1849 on a coral atoll, French Polynesia. *J. Exp. Mar. Biol. Ecol.* 209, 15–31. doi: 10.1016/S0022-0981(96)00057-3

Frank, M. G., Waldrop, R. H., Dumoulin, M., Aton, S., and Boal, J. G. (2012). A preliminary analysis of sleep-like states in the Cuttlefish *Sepia officinalis*. *PLoS ONE* 7, e38125. doi: 10.1371/journal.pone.0038125

Frasnelli, E., Ponte, G., Vallortigara, G., and Fiorito, G. (2019). Visual lateralization in the cephalopod mollusk *Octopus vulgaris*. *Symmetry* 11, 1121. doi: 10.3390/sym11091121

Frith, C. D., and Frith, U. (2006). The neural basis of mentalizing. *Neuron* 50, 531–534. doi: 10.1016/j.neuron.2006.05.001

Gherardi, F., Aquiloni, L., and Tricarico, E. (2012). Revisiting social recognition systems in invertebrates. *Anim. Cogn.* 15, 745–762. doi: 10.1007/s10071-012-0513-y

Gherardi, F., Cenni, F., Parisi, G., and Aquiloni, L. (2010). Visual recognition of conspecifics in the American lobster, *Homarus americanus*. *Anim. Behav.* 80, 713–719. doi: 10.1016/j.anbehav.2010.07.008

Gleadall, I. G. (1990). Higher motor function in the brain of Octopus: the anterior basal lobe and its analogies with the vertebrate basal ganglia. *Ann. Appl. Inf. Sci.* 16, 1–30.

Godfrey-Smith, P. (2013). Cephalopods and the evolution of the mind. *Pacific Conserv. Biol.* 19, 4–9. doi: 10.1071/PC130004

Godfrey-Smith, P. (2016). *Other Minds: The Octopus, the Sea, and the Deep Origins of Consciousness*. London: Farrar, Straus and Giroux.

Godfrey-Smith, P. (2019). Octopus experience. *Anim. Sentience* 4, 18. doi: 10.51291/2377-7478.1503

Godfrey-Smith, P., and Lawrence, M. (2012). Long-term high-density occupation of a site by Octopus tetricus and possible site modification due to foraging behavior. *Mar. Freshw. Behav. Physiol.* 45, 1–8. doi: 10.1080/10236244.2012.727617

Graziadei, P. (1971). "The nervous system of the arm," in *The Anatomy of the Nervous System of Octopus vulgaris*, ed J. Z. Young (London: Oxford University Press), 45–61.

Griffin, D. R. (1976). *The Question of Animal Awareness: Evolutionary Continuity of Mental Experience*. New York, NY: Rockefeller University Press.

Grossmann, T. (2017). The eyes as windows into other minds: an integrative perspective. *Perspect. Psychol. Sci.* 12, 107–121. doi: 10.1177/1745691616654457

Guerra, Á., Hernández-Urcera, J., Garci, M. E., Sestelo, M., Regueira, M., González, Á. F., et al. (2014). Dwellers in dens on sandy bottoms: ecological and behavioural traits of *Octopus vulgaris*. *Sci. Mar.* 78, 405–414. doi: 10.3989/scimar.04071.28F

Gutfreund, Y. (2018). The mind-evolution problem: the difficulty of fitting consciousness in an evolutionary framework. *Front. Psychol.* 9, 1537. doi: 10.3389/fpsyg.2018.01537

Gutfreund, Y. (2019). Who needs a mind when you have thousands of fingers? *Anim. Sentience* 4, 3. doi: 10.51291/2377-7478.1469

Gutnick, T., Byrne, R. A., Hochner, B., and Kuba, M. (2011). *Octopus vulgaris* uses visual information to determine the location of its arm. *Curr. Biol.* 21, 460–462. doi: 10.1016/j.cub.2011.01.052

Gutnick, T., Shomrat, T., Mather, J. A., and Kuba, M. J. (2021). "The cephalopod brain: motion control, learning, and cognition," in *Physiology of Molluscs*, eds S. Saleuddin and S. Mukai (Boca Raton, FL: Apple Academic Press), 137–177. doi: 10.1201/9781315207117-5

Halassa, M. M., Chen, Z., Wimmer, R. D., Brunetti, P. M., Zhao, S., Zikopoulos, B., et al. (2014). State-dependent architecture of thalamic reticular subnetworks. *Cell* 158, 808–821. doi: 10.1016/j.cell.2014.06.025

Halassa, M. M., and Kastner, S. (2017). Thalamic functions in distributed cognitive control. *Nat. Neurosci.* 20, 1669–1679. doi: 10.1038/s41593-017-0020-1

Hanlon, R. T., Forsythe, J. W., and Joneschild, D. E. (1999). Crypsis, conspicuousness, mimicry and polyphenism as antipredator defences of foraging octopuses on Indo-pacific coral reefs, with a method of quantifying crypsis from video tapes. *Biol. J. Linnean Soc.* 66, 1–22. doi: 10.1111/j.1095-8312.1999.tb01914.x

Hanlon, R. T., and Messenger, J. B. (2018). *Cephalopod Behaviour. Cambridge: Cambridge University Press.*

Hepper, P. G. (1986). Kin recognition: functions and mechanisms a review. *Biol. Rev.* 61, 63–93. doi: 10.1111/j.1469-185X.1986.tb00427.x

Hirschfeld, L. A., and Gelman, S. A. (1994). "Toward a topography of mind: an introduction to domain specificity," in *Mapping the Mind: Domain Specificity in Cognition and Culture*, eds L. A. Hirschfeld and S. A. Gelman (Cambridge: Cambridge University Press), 3–35. doi: 10.1017/CBO9780511752902.002

Hochner, B. (2012). An embodied view of octopus neurobiology. *Curr. Biol.* 22, R887–R892. doi: 10.1016/j.cub.2012.09.001

Hochner, B. (2013). How nervous systems evolve in relation to their embodiment: what we can learn from octopuses and other molluscs. *Brain Behav. Evol.* 82, 19–30. doi: 10.1159/000353419

Hochner, B., Brown, E. R., Langella, M., Shomrat, T., and Fiorito, G. (2003). A learning and memory area in the octopus brain manifests a vertebrate-like long-term potentiation. *J. Neurophysiol.* 90, 3547–3554. doi: 10.1152/jn.00645.2003

Hochner, B., and Shomrat, T. (2013). The neurophysiological basis of learning and memory in advanced invertebrates the octopus and the cuttlefish. *Invertebrate Learn. Memory* 22, 303–317. doi: 10.1016/B978-0-12-415823-8.00024-1

Hochner, B., Shomrat, T., and Fiorito, G. (2006). The octopus: a model for a comparative analysis of the evolution of learning and memory mechanisms. *Biol. Bull.* 210, 308–317. doi: 10.2307/4134567

Holden-Dye, L., Ponte, G., Allcock, A. L., Vidal, E. A. G., Nakajima, R., Peterson, T. R., et al. (2019). Editorial: cephsinaction: towards future challenges for cephalopod science. *Front. Physiol.* 10, 980. doi: 10.3389/fphys.2019.00980

Huang, K.-L., and Chiao, C.-C. (2013). Can cuttlefish learn by observing others? *Anim. Cogn.* 16, 313–320. doi: 10.1007/s10071-012-0573-z

Huffard, C. L., and Bartick, M. (2015). Wild Wunderpus photogenicus and *Octopus cyanea* employ asphyxiating "constricting" in interactions with other octopuses. *Molluscan Res.* 35, 12–16. doi: 10.1080/13235818.2014.909558

Huffard, C. L., Caldwell, R. L., and Boneka, F. (2008). Mating behavior of *Abdopus aculeatus* (d'Orbigny 1834) (Cephalopoda: Octopodidae) in the wild. *Mar. Biol.* 154, 353–362. doi: 10.1007/s00227-008-0930-2

Humphrey, N. (2006). *Seeing Red: A Study in Consciousness. New York, NY: Belknap Press; Harvard University Press.*

Iglesias, T. L., Boal, J. G., Frank, M. G., Zeil, J., and Hanlon, R. T. (2019). Cyclic nature of the REM sleep-like state in the cuttlefish *Sepia officinalis*. *J. Exp. Biol.* 222, jeb174862. doi: 10.1242/jeb.174862

James, W. (1977). *The Writings of William James*. A Comprehensive Edition. Chicago, IL: The University of Chicago Press.

Jozet-Alves, C., Bertin, M., and Clayton, N. S. (2013). Evidence of episodic-like memory in cuttlefish. *Curr. Biol.* 23, R1033–R1035. doi: 10.1016/j.cub.2013.10.021

Jozet-Alves, C., Romagny, S., Bellanger, C., and Dickel, L. (2012a). Cerebral correlates of visual lateralization in Sepia. *Behav. Brain Res.* 234, 20–25. doi: 10.1016/j.bbr.2012.05.042

Jozet-Alves, C., Viblanc, V. A., Romagny, S., Dacher, M., Healy, S. D., and Dickel, L. (2012b). Visual lateralization is task and age dependent in cuttlefish, *Sepia officinalis*. *Anim. Behav.* 83, 1313–1318. doi: 10.1016/j.anbehav.2012.02.023

Kano, F., Moore, R., Krupenye, C., Hirata, S., Tomonaga, M., and Call, J. (2018). Human ostensive signals do not enhance gaze following in chimpanzees, but do enhance object-oriented attention. *Anim. Cogn.* 21, 715–728. doi: 10.1007/s10071-018-1205-z

Katz, I., Shomrat, T., and Nesher, N. (2021). Feel the light: sight-independent negative phototactic response in octopus arms. *J. Exp. Biol.* 224, jeb237529. doi: 10.1242/jeb.237529

Kawashima, S., Yasumuro, H., and Ikeda, Y. (2021). Plain-body octopus's (*Callistoctopus aspilosomatis*) learning about objects *via* both visual and tactile sensory inputs: a pilot study. *Zool. Sci.* 38, 383–396. doi: 10.2108/zs210034

Kayes, R. (1974). The daily activity pattern of *Octopus vulgaris* in a natural habitat. *Marine Freshw. Behav. Physiol.* 2, 337–343. doi: 10.1080/10236247309386935

Kuba, M., Meisel, D., Byrne, R., Griebel, U., and Mather, J. (2003). Looking at play in *Octopus vulgaris*. *Berliner Paläontologische Abhandlungen* 3, 163–169.

Kuba, M. J., Byrne, R. A., Meisel, D. V., and Mather, J. A. (2006). When do octopuses play? *Effects of repeated testing, object type, age, and food deprivation on object play in Octopus vulgaris. J. Compar. Psychol.* 120, 184. doi: 10.1037/0735-7036.120.3.184

Kuo, T.-H., and Chiao, C.-C. (2020). Learned valuation during forage decision-making in cuttlefish. *R. Soc. Open Sci.* 7, 201602. doi: 10.1098/rsos.201602

Lane, F. W. (1960). *Kingdom of the Octopus; the Life History of the Cephalopoda*. New York, NY: Sheridan House.

Lee, H. (1875). The Octopus: Or, the "Devil-Fish" of Fiction and of Fact. *London: Chapman and Hall.*

Maddock, L., and Young, J. Z. (1987). Quantitative differences among the brains of cephalopods. *J. Zool.* 212, 739–767. doi: 10.1111/j.1469-7998.1987.tb05967.x

Makalic, E. (2010). *Hypothesis Testing With Paul the Octopus*. Available online at: http://www.emakalic.org/blog/?p=40

Maldonado, H. (1963a). The general amplification function of the vertical lobe in *Octopus vulgaris*. *J. Compar. Physiol. A Neuroethol. Sensory Neural Behav. Physiol.* 47, 215–229. doi: 10.1007/BF00298034

Maldonado, H. (1963b). The positive learning process in *Octopus vulgaris*. *Zeitschrift vergleichende Physiol.* 47, 191–214. doi: 10.1007/BF00303120

Maldonado, H. (1963c). The visual attack learning system in *Octopus vulgaris*. *J. Theor. Biol.* 5, 470–488. doi: 10.1016/0022-5193(63)90090-0

Maldonado, H. (1965). The positive and negative learning process in *Octopus vulgaris* Lamarck. *Influence of the vertical and median superior frontal lobes. Zeitschrift vergleichende Physiologie* 51, 185–203. doi: 10.1007/BF00299293

Maldonado, H. (1968). Effect of electroconvulsive shock on memory in *Octopus vulgaris* Lamarck. *Z. vergl. Physiol.* 59, 25–37. doi: 10.1007/BF00298809

Maldonado, H. (1969). Futher investigations on the effect of electroconvulsive shock (ECS) on memory in *Octopus vulgaris*. *Z. vergl. Physiol.* 63, 113–118. doi: 10.1007/BF00298333

Manzo, P. (2021). *Learning and memory in Octopus vulgaris: search of the underlying biological machinery* (PhD). Università della Calabria, Rende, Italy.

Marini, G., De Sio, F., Ponte, G., and Fiorito, G. (2017). "Behavioral analysis of learning and memory in cephalopods," in *Learning and Memory: A Comprehensive Reference, 2nd Edn.*, ed J. H. Byrne (Amsterdam: Academic Press, Elsevier), 441–462.

Masciari, C. F., and Carruthers, P. (2021). Perceptual awareness or phenomenal consciousness? A dilemma. *Biol. Philos.* 36, 18. doi: 10.1007/s10539-021-09795-1

Maselli, V., Al-Soudy, A.-S., Buglione, M., Aria, M., Polese, G., and Di Cosmo, A. (2020). Sensorial hierarchy in *Octopus vulgaris*'s food choice: chemical vs. *visual. Animals* 10, 457. doi: 10.3390/ani10030457

Mather, J. (2019). What is in an octopus' mind? *Anim. Sentience* 26, 1–29. doi: 10.51291/2377-7478.1370

Mather, J. (2021a). The case for octopus consciousness: unity. *NeuroScience* 2, 405–415. doi: 10.3390/neurosci2040030

Mather, J. (2021b). Octopus consciousness: the role of perceptual richness. *NeuroScience* 2, 276–290. doi: 10.3390/neurosci2030020

Mather, J. A. (1991). Navigation by spatial memory and use of visual landmarks in octopuses. *J. Comp. Physiol. A* 168, 491–497. doi: 10.1007/BF00199609

Mather, J. A. (2008). Cephalopod consciousness: behavioural evidence. *Conscious. Cogn.* 17, 37–48. doi: 10.1016/j.concog.2006.11.006

Mather, J. A., and Anderson, R. C. (1993). Personalities of octopuses (*Octopus rubescens*). *J. Comp. Psychol.* 107, 336–340. doi: 10.1037/0735-7036.107.3.336

Mather, J. A., and Anderson, R. C. (1999). Exploration, play and habituation in octopuses (*Octopus dofleini*). *J. Comp. Psychol.* 113, 333. doi: 10.1037/0735-7036.113.3.333

Mather, J. A., Anderson, R. C., and Wood, J. B. (2010). *Octopus: The Ocean's Intelligent Invertebrate*. Portland, OR: Timber Press.

Mather, J. A., and Dickel, L. (2017). Cephalopod complex cognition. *Curr. Opin. Behav. Sci.* 16, 131–137. doi: 10.1016/j.cobeha.2017.06.008

Mather, J. A., and O'Dor, R. K. (1991). Foraging strategies and predation risk shape the natural history of juvenile *Octopus vulgaris*. *Bull. Mar. Sci.* 49, 256–269.

Mäthger, L. M., and Hanlon, R. T. (2006). Anatomical basis for camouflaged polarized light communication in squid. *Biol. Lett.* 2, 494–496. doi: 10.1098/rsbl.2006.0542

Medeiros, S. L. d. S., Paiva, M. M. M. d., Lopes, P. H., Blanco, W., Lima, F. D. d., Oliveira, J. B. C. d., et al. (2021). Cyclic alternation of quiet and active sleep states in the octopus. *iScience* 24, 102223. doi: 10.1016/j.isci.2021.102223

Meisel, D. V., Byrne, R. A., Mather, J. A., and Kuba, M. (2011). Behavioral sleep in *Octopus vulgaris*. *Vie et Milieu* 61, 185–190.

Messenger, J. B. (1973). Learning in the cuttlefish, Sepia. *Anim. Behav.* 21, 801–826. doi: 10.1016/S0003-3472(73)80107-1

Messenger, J. B. (1996). Neurotransmitters of cephalopods. *Invertebr. Neurosci* 2, 95–114. doi: 10.1007/BF02214113

Mezrai, N., Chiao, C.-C., Dickel, L., and Darmaillacq, A.-S. (2019). A difference in timing for the onset of visual and chemosensory systems during embryonic development in two closely related cuttlefish species. *Dev. Psychobiol.* 61, 1014–1021. doi: 10.1002/dev.21868

Montgomery, S. (2015). *The Octopus Scientists: Exploring the Mind of a Mollusk*. Boston, MA: Houghton Mifflin Harcourt.

Morse, P., and Huffard, C. L. (2019). Tactical tentacles: new insights on the processes of sexual selection among the cephalopoda. *Front. Physiol.* 10, 1035. doi: 10.3389/fphys.2019.01035

Morse, P., Zenger, K. R., McCormick, M. I., Meekan, M. G., and Huffard, C. L. (2017). Chemical cues correlate with agonistic behaviour and female mate choice in the southern blue-ringed octopus, *Hapalochlaena maculosa* (Hoyle, 1883) (Cephalopoda: Octopodidae). *J. Molluscan Stud.* 83, 79–87. doi: 10.1093/mollus/eyw045

Müller, C. A., Schmitt, K., Barber, A. L., and Huber, L. (2015). Dogs can discriminate emotional expressions of human faces. *Curr. Biol.* 25, 601–605. doi: 10.1016/j.cub.2014.12.055

Myers, C. E. (1992). *Delay Learning in Artificial Neural Networks*. London: Chapman & Hall.

Nagasawa, M., Murai, K., Mogi, K., and Kikusui, T. (2011). Dogs can discriminate human smiling faces from blank expressions. *Anim. Cogn.* 14, 525–533. doi: 10.1007/s10071-011-0386-5

Nakajima, R. (2018). "Can I talk to a squid? The origin of visual communication through the behavioral ecology of Cephalopod," in *Human Interface and the Management of Information. Interaction, Visualization, and Analytics*, eds S. Yamamoto and H. Mori (Cham: Springer International Publishing), 594–606.

Nakajima, R., Shigeno, S., Zullo, L., De Sio, F., and Schmidt, M. R. (2018). Cephalopods between science, art, and engineering: a contemporary synthesis. *Front. Commun.* 3, 20. doi: 10.3389/fcomm.2018.00020

Nawroth, C., Trincas, E., and Favaro, L. (2017). African penguins follow the gaze direction of conspecifics. *PeerJ* 5, e3459. doi: 10.7717/peerj.3459

Nesher, N., Levy, G., Grasso, F. W., and Hochner, B. (2014). Self-recognition mechanism between skin and suckers prevents octopus arms from interfering with each other. *Curr. Biol.* 24, 1271–1275. doi: 10.1016/j.cub.2014.04.024

Nixon, M., and Young, J. Z. (2003). *The Brains and Lives of Cephalopods*. New York, NY: Oxford University.

Norman, E. (2017). Metacognition and mindfulness: the role of fringe consciousness. *Mindfulness* 8, 95–100. doi: 10.1007/s12671-016-0494-z

O'Brien, C. E., Di Miccoli, V., and Fiorito, G. (2021). A preliminary investigation of the response of *Octopus vulgaris* to experimental stimuli in the wild. *J. Molluscan Stud.* 87, eyab032. doi: 10.1093/mollus/eyab032

Oshima, M., di Pauli von Treuheim, T., Carroll, J., Hanlon, R. T., Walters, E. T., and Crook, R. J. (2016). Peripheral injury alters schooling behavior in squid, *Doryteuthis pealeii*. *Behav. Processes* 128, 89–95. doi: 10.1016/j.beproc.2016.04.008

Owen, R. (1832). Memoir on the Pearly Nautilus (Nautilus Pompilius, Linn.). *London: Richard Taylor*.

Packard, A. (1963). The behaviour of *Octopus vulgaris*. *Bull. l'Institut océanographique (Monaco) Numéro spécial* 1D, 35–49.

Packard, A. (1972). Cephalopods and fish: the limits of convergence. *Biol. Rev.* 47, 241–307. doi: 10.1111/j.1469-185X.1972.tb00975.x

Packard, A., and Hochberg, F. G. (1977). Skin patterning in Octopus and other Genera. *Symposia Zool. Soc. London* 38, 191–231.

Packard, A., and Sanders, G. D. (1971). Body patterns of *Octopus vulgaris* and maturation of the response to disturbance. *Anim. Behav.* 19, 780–790. doi: 10.1016/S0003-3472(71)80181-1

Partan, S. R., and Marler, P. (2005). Issues in the classification of multimodal communication signals. *Am. Nat.* 166, 231–245. doi: 10.1086/431246

Pinto-Bazurco, J. F. (2020). *The Precautionary Principle*. International Institute for Sustainable Development (IISD).

Plän, T. (1987). *Funktionelle Neuroanatomie sensorisch/motorischer loben im gehirn von Octopus vulgaris*. Doktorgrades der Naturwissehschaften (Dr. Rer. Nat.), Regensburg: Universität Regensburg.

Pliny, t. E. (1961). *Naturalis Historia, with an English translation by H. Rackham*. Cambridge, MA: Harvard University Press.

Polese, G., Bertapelle, C., and Di Cosmo, A. (2015). Role of olfaction in *Octopus vulgaris* reproduction. *Gen. Comp. Endocrinol.* 210, 55–62. doi: 10.1016/j.ygcen.2014.10.006

Poncet, L., Roig, A., Billard, P., Bellanger, C., and Jozet-Alves, C. (2020). *Future Planning Abilities in the Common Cuttlefish. CephRes2020 Virtual Event, Sep 2020, Napoli, Italy*. Napoli: CephRes Reference Docs.

Ponte, G. (2012). *Distribution and preliminary functional analysis of some modulators in the cephalopod mollusc Octopus vulgaris* (PhD Thesis). Università della Calabria, Italy; Stazione Zoologica Anton Dohrn, Napoli, Italy.

Ponte, G., and Fiorito, G. (2015). "Immunohistochemical Analysis of Neuronal Networks in the Nervous System of *Octopus vulgaris*," in *Immunocytochemistry and Related Techniques,* eds A. Merighi and L. Lossi., Neuromethods 101, 61–77. doi: 10.1007/978-1-4939-2313-7_3

Ponte, G., Taite, M., Borrelli, L., Tarallo, A., Allcock, A. L., and Fiorito, G. (2021). Cerebrotypes in cephalopods: brain diversity and its correlation with species habits, life history, and physiological adaptations. *Front. Neuroanat.* 14, 565109. doi: 10.3389/fnana.2020.565109

Power, J. (1857). Observations on the habits of various marine animals. *Observations upon Octopus vulgaris and Pinna nobilis. Ann. Magazine Natural Hist.* 20, 336. doi: 10.1080/00222935709487931

Premack, D. (1988). "'Does the chimpanzee have a theory of mind'revisited," in *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans*, eds R. W. Byrne and A. Whiten (New York, NY: Clarendon Press; Oxford University Press), 160–179.

Premack, D., and Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behav. Brain Sci.* 1, 515–526. doi: 10.1017/S0140525X00076512

Pronk, R., Wilson, D., and Harcourt, R. (2010). Video playback demonstrates episodic personality in the gloomy octopus. *J. Exp. Biol.* 213, 1035–1041. doi: 10.1242/jeb.040675

Purdy, J. E., Dixon, D., Estrada, A., Peters, A., Riedlinger, E., and Suarez, R. (2006). Prawn-in-a-tube procedure: habituation or associative learning in cuttlefish? *J. Gen. Psychol.* 133, 131–152. doi: 10.3200/GENP.133.2.131-152

Racca, A., Amadei, E., Ligout, S., Guo, K., Meints, K., and Mills, D. (2010). Discrimination of human and dog faces and inversion responses in domestic dogs (*Canis familiaris*). *Anim. Cogn.* 13, 525–533. doi: 10.1007/s10071-009-0303-3

Rajneesh, K., and Bolash, R. (2018). "Pathways of pain perception and modulation," in *Fundamentals of Pain Medicine*, eds J. Cheng and R. W. Rosenquist (Cham: Springer International Publishing), 7–11.

Rankin, C. H. (2004). Invertebrate learning: what can't a worm learn? *Curr. Biol.* 14, R617–R618. doi: 10.1016/j.cub.2004.07.044

Redinbaugh, M. J., Phillips, J. M., Kambi, N. A., Mohanta, S., Andryk, S., Dooley, G. L., et al. (2020). Thalamus modulates consciousness *via* layer-specific control of cortex. *Neuron 106,* 66–75.e12. doi: 10.1016/j.neuron.2020.01.005

Reiter, S., Hülsdunk, P., Woo, T., Lauterbach, M. A., Eberle, J. S., Akay, L. A., et al. (2018). Elucidating the control and development of skin patterning in cuttlefish. *Nature* 562, 361–366. doi: 10.1038/s41586-018-0591-3

Reiter, S., and Laurent, G. (2020). Visual perception and cuttlefish camouflage. *Curr. Opin. Neurobiol.* 60, 47–54. doi: 10.1016/j.conb.2019.10.010

Romanes, G. J. (1885). *Mental Evolution in Animals*. London: Kegan Paul, Trench & Co.

Roth, G. (2015). Convergent evolution of complex brains and high intelligence. *Philos. Trans. R. Soc. B Biol. Sci.* 370, 1684. doi: 10.1098/rstb.2015.0049

Rowe, C., and Guilford, T. (1999). The evolution of multimodal warning displays. *Evol. Ecol.* 13, 655–671. doi: 10.1023/A:1011021630244

Rybarczyk, P., Koba, Y., Rushen, J., Tanida, H., and de Passillé, A. M. (2001). Can cows discriminate people by their faces? *Appl. Anim. Behav. Sci.* 74, 175–189. doi: 10.1016/S0168-1591(01)00162-9

Ryle, G. (1949). *The Concept of Mind*. London: Hutchinson-s Univeristy Library.

Sanders, F. K., and Young, J. Z. (1940). Learning and other functions of the higher nervous centres of Sepia. *J. Neurophysiol.* 3, 501–526. doi: 10.1152/jn.1940.3.6.501

Sanders, G. D. (1975). "The cephalopods," in *Invertebrate Learning. Cephalopods and Echinoderms*, eds W. C. Corning, J. A. Dyal, and A. O. D. Willows (New York, NY: Plenum Press), 1–101.

Scatà, G., Darmaillacq, A. S., Dickel, L., McCusker, S., and Shashar, N. (2017). Going up or sideways? Perception of space and obstacles negotiating by cuttlefish. *Front. Physiol.* 8, 173. doi: 10.3389/fphys.2017.00173

Scheel, D., Godfrey-Smith, P., and Lawrence, M. (2016). Signal use by octopuses in agonistic interactions. *Curr. Biol.* 26, 377–382. doi: 10.1016/j.cub.2015.12.033

Scheel, D., Leite, T., Mather, J., and Langford, K. (2017). Diversity in the diet of the predator *Octopus cyanea* in the coral reef system of Moorea, French Polynesia. *J. Nat. Hist.* 51, 2615–2633. doi: 10.1080/00222933.2016.1244298

Schiff, N. D. (2008). Central thalamic contributions to arousal regulation and neurological disorders of consciousness. *Ann. N. Y. Acad. Sci.* 1129, 105–118. doi: 10.1196/annals.1417.029

Schmitt, L. I., Wimmer, R. D., Nakajima, M., Happ, M., Mofakham, S., and Halassa, M. M. (2017). Thalamic amplification of cortical connectivity sustains attentional control. *Nature* 545, 219–223. doi: 10.1038/nature22073

Schneider, G. H. (1880). *Der thierische Wille, systematische Darstellung und Erklärung der thierischen Triebe und deren Entstehung, Entwickelung und Verbreitung im Thierreiche als Grundlage zu einer vergleichenden Willenslehre*. Leipzig: Abel.

Schnell, A. K., Amodio, P., Boeckle, M., and Clayton, N. S. (2021c). How intelligent is a cephalopod? Lessons from comparative cognition. *Biol. Rev.* 96, 162–178. doi: 10.1111/brv.12651

Schnell, A. K., Bellanger, C., Vallortigara, G., and Jozet-Alves, C. (2018). Visual asymmetries in cuttlefish during brightness matching for camouflage. *Curr. Biol.* 28, R925–R926. doi: 10.1016/j.cub.2018.07.019

Schnell, A. K., Boeckle, M., Rivera, M., Clayton, N. S., and Hanlon, R. T. (2021a). Cuttlefish exert self-control in a delay of gratification task. *Proc. R. Soc. B Biol. Sci.* 288, 20203161. doi: 10.1098/rspb.2020.3161

Schnell, A. K., Clayton, N. S., Hanlon, R. T., and Jozet-Alves, C. (2021b). Episodic-like memory is preserved with age in cuttlefish. *Proc. R. Soc. B Biol. Sci.* 288, 20211052. doi: 10.1098/rspb.2021.1052

Schnell, A. K., Hanlon, R. T., Benkada, A., and Jozet-Alves, C. (2016a). Lateralization of eye use in cuttlefish: opposite direction for anti-predatory and predatory behaviors. *Front. Physiol.* 7, 620. doi: 10.3389/fphys.2016.00620

Schnell, A. K., Smith, C. L., Hanlon, R. T., Hall, K. C., and Harcourt, R. (2016b). Cuttlefish perform multiple agonistic displays to communicate a hierarchy of threats. *Behav. Ecol. Sociobiol.* 70, 1643–1655. doi: 10.1007/s00265-016-2170-7

Schwartz, B. L. (2019). A community of minds. *Anim. Sentience* 26, 2. doi: 10.51291/2377-7478.1468

Schweid, R. (2013). *Octopus*. London: Reaktion Books.

Seager, W. (2016). Theories of Consciousness: An Introduction and Assessment. *New York, NY: Routledge*.

Seth, A. K., Baars, B. J., and Edelman, D. B. (2005). Criteria for consciousness in humans and other mammals. *Conscious. Cogn.* 14, 119–139. doi: 10.1016/j.concog.2004.08.006

Shashar, N., Vaughan, K., Loew, E., Boal, J., Hanlon, R., and Grable, M. (2004). Behavioral evidence for intraspecific signaling with achromatic and polarized light by cuttlefish (Mollusca: Cephalopoda). *Behaviour* 141, 837–861. doi: 10.1163/1568539042265662

Sheehan, M. J., and Tibbetts, E. A. (2011). Specialized face learning is associated with individual recognition in paper wasps. *Science* 334, 1272–1275. doi: 10.1126/science.1211334

Shigeno, S., Andrews, P. L. R., Ponte, G., and Fiorito, G. (2018). Cephalopod brains: an overview of current knowledge to facilitate comparison with vertebrates. *Front. Physiol.* 9, 952. doi: 10.3389/fphys.2018.00952

Shigeno, S., Parnaik, R., Albertin, C. B., and Ragsdale, C. W. (2015). Evidence for a cordal, not ganglionic, pattern of cephalopod brain neurogenesis. *Zool. Lett.* 1, 26. doi: 10.1186/s40851-015-0026-z

Shigeno, S., and Ragsdale, C. W. (2015). The gyri of the octopus vertical lobe have distinct neurochemical identities. *J. Compar. Neurol.* 523, 1297–1317. doi: 10.1002/cne.23755

Shomrat, T., Graindorge, N., Bellanger, C., Fiorito, G., Loewenstein, Y., and Hochner, B. (2011). Alternative sites of synaptic plasticity in two homologous "fan-out fan-in" learning and memory networks. *Curr. Biol.* 21, 1773–1782. doi: 10.1016/j.cub.2011.09.011

Shomrat, T., Turchetti-Maia, A. L., Stern-Mentch, N., Basil, J. A., and Hochner, B. (2015). The vertical lobe of cephalopods: an attractive brain structure for understanding the evolution of advanced learning and memory systems. *J. Compar. Physiol. A* 201, 947–956. doi: 10.1007/s00359-015-1023-6

Shomrat, T., Zarrella, I., Fiorito, G., and Hochner, B. (2008). The octopus vertical lobe modulates short-term learning rate and uses LTP to acquire long-term memory. *Curr. Biol.* 18, 337–342. doi: 10.1016/j.cub.2008.01.056

Sih, A., Bell, A., and Johnson, J. C. (2004a). Behavioral syndromes: an ecological and evolutionary overview. *Trends Ecol. Evol* 19, 372–378. doi: 10.1016/j.tree.2004.04.009

Sih, A., Bell, A. M., Johnson, J. C., and Ziemba, R. E. (2004b). Behavioral syndromes: an integrative overview. *Q. Rev. Biol.* 79, 241–277. doi: 10.1086/422893

Sinn, D. L., Gosling, S. D., and Moltschaniwskyj, N. A. (2008). Development of shy/bold behaviour in squid: context-specific phenotypes associated with developmental plasticity. *Anim. Behav.* 75, 433–442. doi: 10.1016/j.anbehav.2007.05.008

Sinn, D. L., Moltschaniwskyj, N. A., Wapstra, E., and Dall, S. R. X. (2010). Are behavioral syndromes invariant? *Spatiotemporal variation in shy/bold behavior in squid. Behav. Ecol. Sociobiol.* 64, 693–702. doi: 10.1007/s00265-009-0887-2

Sinn, D. L., Perrin, N. A., Mather, J. A., and Anderson, R. C. (2001). Early temperamental traits in an octopus (*Octopus bimaculoides*). *J. Comp. Psychol* 115, 351–364. doi: 10.1037/0735-7036.115.4.351

Smith, J. A., Andrews, P. L. R., Hawkins, P., Louhimies, S., Ponte, G., and Dickel, L. (2013). Cephalopod research and EU Directive 2010/63/EU: requirements, impacts and ethical review. *J. Exp. Mar. Biol. Ecol.* 447, 31–45. doi: 10.1016/j.jembe.2013.02.009

Snijders, L., and Naguib, M. (2017). Communication in animal social networks: a missing link? *Adv. Study Behav*. 2017, 297–359. doi: 10.1016/bs.asb.2017.02.004

Steele, R. E. (2020). Gene editing: a tool for tackling cephalopod biology. *Curr. Biol.* 30, R986–R988. doi: 10.1016/j.cub.2020.06.094

Stone, S. M. (2010). Human facial discrimination in horses: can they tell us apart? *Anim. Cogn.* 13, 51–61. doi: 10.1007/s10071-009-0244-x

Sumbre, G., Fiorito, G., Flash, T., and Hochner, B. (2005). Neurobiology: motor control of flexible octopus arms. *Nature* 433, 595. doi: 10.1038/433595a

Sumbre, G., Fiorito, G., Flash, T., and Hochner, B. (2006). Octopuses use a human-like strategy to control precise point-to-point arm movements. *Curr. Biol.* 16, 767–772. doi: 10.1016/j.cub.2006.02.069

Sumbre, G., Gutfreund, Y., Fiorito, G., Flash, T., and Hochner, B. (2001). Control of octopus arm extension by a peripheral motor program. *Science* 293, 1845–1848. doi: 10.1126/science.1060976

Swanson, L. W. (2007). Quest for the basic plan of nervous system circuitry. *Brain Res. Rev.* 55, 356–372. doi: 10.1016/j.brainresrev.2006.12.006

Tanida, H., and Nagano, Y. (1998). The ability of miniature pigs to discriminate between a stranger and their familiar handler. *Appl. Anim. Behav. Sci.* 56, 149–159. doi: 10.1016/S0168-1591(97)00095-6

Tessmar-Raible, K. (2007). The evolution of neurosecretory centers in bilaterian forebrains: insights from protostomes. *Semin. Cell Dev. Biol.* 18, 492–501. doi: 10.1016/j.semcdb.2007.04.007

Tessmar-Raible, K., Raible, F., Christodoulou, F., Guy, K., Rembold, M., Hausen, H., et al. (2007). Conserved sensory-neurosecretory cell types in annelid and fish forebrain: insights into hypothalamus evolution. *Cell* 129, 1389–1400. doi: 10.1016/j.cell.2007.04.041

Tibbetts, E. A. (2002). Visual signals of individual identity in the wasp Polistes fuscatus. *Proc. R. Soc. London B Biol. Sci.* 269, 1423–1428. doi: 10.1098/rspb.2002.2031

Tibbetts, E. A., and Dale, J. (2007). Individual recognition: it is good to be different. *Trends Ecol. Evol.* 22, 529–537. doi: 10.1016/j.tree.2007.09.001

Tomita, M., and Aoki, S. (2014). Visual discrimination learning in the small octopus *Octopus ocellatus*. *Ethology* 120, 863–872. doi: 10.1111/eth.12258

Tricarico, E., Amodio, P., Ponte, G., and Fiorito, G. (2014). "Cognition and recognition in the cephalopod mollusc *Octopus vulgaris*: coordinating interaction with environment and conspecifics," in *Biocommunication of Animals*, ed G. Witzany (Dordrecht: Springer Science+Business Media), 337–349.

Tricarico, E., Borrelli, L., Gherardi, F., and Fiorito, G. (2011). I know my neighbour: individual recognition in *Octopus vulgaris*. *PLoS ONE* 6, e0018710. doi: 10.1371/journal.pone.0018710

Turchetti-Maia, A., Shomrat, T., and Hochner, B. (2017). "The vertical lobe of cephalopods. a brain structure ideal for exploring the mechanisms of complex forms of learning and memory," in *The Oxford Handbook of Invertebrate Neurobiology*, ed J. J. Byrne (Oxford: Oxford University Press), 1–27.

Vallortigara, G. (2017). Sentience does not require "higher" cognition. *Anim. Sentience* 2, 6. doi: 10.51291/2377-7478.1226

Van der Velden, J., Zheng, Y., Patullo, B. W., and Macmillan, D. L. (2008). Crayfish recognize the faces of fight opponents. *PLoS ONE* 3, e1695. doi: 10.1371/journal.pone.0001695

van Giesen, L., Kilian, P. B., Allard, C. A. H., and Bellono, N. W. (2020). Molecular basis of chemotactile sensation in octopus. *Cell 183,* 594–604.e514. doi: 10.1016/j.cell.2020.09.008

van Woerkum, B. (2020). Distributed nervous system, disunified consciousness?: A sensorimotor integrationist account of octopus consciousness. *J. Consciousness Stud.* 27, 149–172.

Villanueva, R., Perricone, V., and Fiorito, G. (2017). Cephalopods as predators: a short journey among behavioral flexibilities, adaptions, and feeding habits. *Front. Physiol.* 8, 598. doi: 10.3389/fphys.2017.00598

Vitti, J. (2010). *The Distribution and Evolution of Animal Consciousness* (Bachelor of Arts with Honors in Philosophy), Harvard University, Cambridge, MA, United States.

Wells, M. J. (1960). Proprioception and visual discrimination of orientation in Octopus. *J. Exp. Biol.* 37, 489–499. doi: 10.1242/jeb.37.3.489

Wells, M. J. (1978). Octopus: physiology and behaviour of an advanced invertebrate. *Springer Science & Business Media*. doi: 10.1007/978-94-017-2468-5

Wilkinson, A., Mandl, I., Bugnyar, T., and Huber, L. (2010). Gaze following in the red-footed tortoise (Geochelone carbonaria). *Anim. Cogn.* 13, 765–769. doi: 10.1007/s10071-010-0320-2

Wilson, A. C. (1985). The molecular basis of evolution. *Sci. Am.* 253, 148–157. doi: 10.1038/scientificamerican1085-164

Wilson, D. S., Clark, A. B., Coleman, K., and Dearstyne, T. (1994). Shyness and boldness in humans and other animals. *Trends Ecol. Evol* 9, 442–446. doi: 10.1016/0169-5347(94)90134-1

Wood, S. M., and Wood, J. N. (2015). Face recognition in newly hatched chicks at the onset of vision. *J. Exp. Psychol. Anim. Learn. Cogn.* 41, 206. doi: 10.1037/xan0000059

Yarnall, J. L. (1969). Aspects of the behaviour of *Octopus cyanea* Gray. *Anim. Behav.* 17, 747–754. doi: 10.1016/S0003-3472(69)80022-9

Young, J. (1954). "Memory, heredity and information," in *Evolution as a Process,* 281.

Young, J. Z. (1951). *Doubt and Certainty in Science: A Biologist's Reflections on the Brain*. Oxford: Clarendon Press.

Young, J. Z. (1961). Learning and discrimination in the octopus. *Biol. Rev.* 36, 32–96. doi: 10.1111/j.1469-185X.1961.tb01432.x

Young, J. Z. (1963). The number and sizes of nerve cells in Octopus. *Proc. Zool. Soc. Lond.* 140, 229–254. doi: 10.1111/j.1469-7998.1963.tb01862.x

Young, J. Z. (1964). *A Model of the Brain*. Oxford: Clarendon Press.

Young, J. Z. (1965a). The buccal nervous system of Octopus. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 249, 27–44. doi: 10.1098/rstb.1965.0007

Young, J. Z. (1965b). The central nervous system of Nautilus. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 249, 1–25. doi: 10.1098/rstb.1965.0006

Young, J. Z. (1967). The visceral nerves of Octopus. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 253, 1–22. doi: 10.1098/rstb.1967.0032

Young, J. Z. (1970). Neurovenous tissues in cephalopods. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 257, 309–321. doi: 10.1098/rstb.1970.0027

Young, J. Z. (1971). *The anatomy of the Nervous System of Octopus vulgaris*. London: Oxford University Press.

Young, J. Z. (1977a). Brain, behaviour and evolution of cephalopods. *Symp. Zool. Soc. Lond*. 38, 377–434.

Young, J. Z. (1977b). The nervous system of Loligo III. *Higher motor centres: the basal supraoesophageal lobes. Philos. Trans. R. Soc. Lond. B* 276, 351–398. doi: 10.1098/rstb.1977.0003

Young, J. Z. (1979). The nervous system of Loligo: V. *The vertical lobe complex. Philos. Trans. R. Soc. Lond. B Biol. Sci.* 285, 311–354. doi: 10.1098/rstb.1979.0008

Young, J. Z. (1991). Computation in the learning system of cephalopods. *Biol. Bull.* 180, 200–208. doi: 10.2307/1542389

Young, J. Z. (1995). "Multiple matrices in the memory system of Octopus," in *Cephalopod Neurobiology*, eds J. N. Abbott, R. Williamson, and L. Maddock (Oxford: Oxford University Press), 431–443.

Zarrella, I. (2011). *Testing changes in gene expression profiles for Octopus vulgaris (Mollusca*, Cephalopoda) (PhD). Napoli: Stazione Zoologica Anton Dohrn, Italy; Open University United Kingdom.

Zarrella, I., Ponte, G., Baldascino, E., and Fiorito, G. (2015). Learning and memory in *Octopus vulgaris*: a case of biological plasticity. *Curr. Opin. Neurobiol.* 35, 74–79. doi: 10.1016/j.conb.2015.06.012

Zoratto, F., Cordeschi, G., Grignani, G., Bonanni, R., Alleva, E., Nascetti, G., et al. (2018). Variability in the "stereotyped" prey capture sequence of male cuttlefish (*Sepia officinalis*) could relate to personality differences. *Anim. Cogn.* 21, 773–785. doi: 10.1007/s10071-018-1209-8

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Systems Neuroscience

TYPE Hypothesis and Theory
PUBLISHED 10 August 2022
DOI 10.3389/fnsys.2022.945722

OPEN ACCESS

EDITED BY
Louis Neal Irwin,
The University of Texas at El Paso,
United States

REVIEWED BY
Jon Mallatt,
Washington State University,
United States
Giorgio Marchetti,
Mind, Consciousness and Language
Research Center, Italy

*CORRESPONDENCE
Thurston Lacalli
lacalli@uvic.ca

RECEIVED 16 May 2022
ACCEPTED 11 July 2022
PUBLISHED 10 August 2022

CITATION
Lacalli T (2022) On the origins
and evolution of qualia: An
experience-space perspective.
Front. Syst. Neurosci. 16:945722.
doi: 10.3389/fnsys.2022.945722

On the origins and evolution of qualia: An experience-space perspective

Thurston Lacalli*

Department of Biology, University of Victoria, Victoria, BC, Canada

This paper elaborates on a proposal for mapping a configuration space for selector circuits (SCs), defined as the subset of neural correlates of consciousness (NCCs) responsible for evoking particular qualia, to its experiential counterpart, experience-space (E-space), as part of an investigation into the nature of conscious experience as it first emerged in evolution. The dimensionality of E-space, meaning the degrees of freedom required to specify the properties of related sets of qualia, is at least two, but the utility of E-space as a hypothetical construct is much enhanced by assuming it is a large dimensional space, with at least several times as many dimensions as there are categories of qualia to occupy them. Phenomenal consciousness can then be represented as having originated as one or more multidimensional ur-experiences that combined multiple forms of experience together. Taking this as a starting point, questions concerning evolutionary sequence can be addressed, including how the quale best suited to a given sensory modality would have been extracted by evolution from a larger set of possibilities, a process referred to here as dimensional sorting, and how phenomenal consciousness would have been experienced in its earliest manifestations. There is a further question as to whether the E-space formulation is meaningful in analytical terms or simply a descriptive device in graphical form, but in either case it provides a more systematic way of thinking about early stages in the evolution of consciousness than relying on narrative and conjecture alone.

KEYWORDS

qualia, phenomenal experience, evolution of consciousness, E-space, dimensional sorting

Introduction

Much of the explanatory success of the scientific enterprise flows from the power of the reductionist enterprise, where a phenomenon is understood by investigating the structure and dynamics of subcomponents of which it is constructed. This methodology has long since proven its utility where those subcomponents have a material existence

and behave in ways that can be observed and measured, whether stars and planets or atoms and quarks. It is problematic when we come to investigate consciousness, whose subcomponents, the contents of consciousness, are neither material in nature nor assignable to a specific spatial location. The most intractable issues, the hard problems of consciousness, relate to the nature of phenomenal experience and its physical source (Levine, 1983, 2009; Chalmers, 1995). However, from a developmental perspective, there is a more prosaic problem of explaining how the neural circuits responsible for generating and/or evoking such experiences are correctly assembled in the embryo. I examined this issue in a preliminary way in an earlier paper (Lacalli, 2020) that explored how Alan Turing's ideas about the emergence of pattern during development might be applied to explain the emergence of consciousness during evolution. Only questions concerning weak emergence (*sensu* Bedau, 1997) can be addressed by this means, which restricts the analysis to the proximate physical correlates and determinants of subjective experience (here, by convention, simply "experience"), meaning the assembly of the relevant neural circuitry. The problem of emergence at the material level is then solved, at least in principle: that given the random variations in circuitry and neural activity that inevitably arise during brain development, the reordering required for consciousness to emerge from the preconscious condition is a matter of having mechanisms in place to selectively amplify those few variants that incrementally move the system toward consciousness. The process as a whole can be characterized as the extraction of order from fluctuations across time scales, because amplification occurs both in real time during development, and across evolutionary time through changes in gene frequencies.

The analysis was extended in a second paper (Lacalli, 2021) on a specific subset of neural correlates of consciousness, namely the selector circuits (SCs) responsible for evoking a particular experience rather than some other, to better understand how SCs behave in response to natural selection. SCs are equivalent in this usage to difference makers of consciousness (DMCs, Klein et al., 2020; see also Hohwy and Bayne, 2015), and are less neutral in a causal sense than the broader category of NCCs (Neisser, 2012), that is, they are more than just correlates. And, it should be pointed out, that so long as consciousness is assumed to be a consequence of neural activity, the DMC/SC formulation is valid regardless of what theory of consciousness one adopts. That is, even for higher order theories that take a representational view of consciousness, that it resides in the algorithmic processing of neural input in and of itself (Van Gulick, 2018; Lycan, 2019; Seth and Bayne, 2022), there will necessarily be components of brain circuitry, whether localized or distributed diffusely across cortical networks, that govern the precise form of experience evoked by a particular sensory input. A configuration space representation is then a useful way of exploring how the constraints on SCs for the simplest of conscious contents change over evolutionary time. How a configurational, neurocircuitry-based SC-space might map to an experiential space (E-space) is a separate issue, and there are no clear guidelines as to how best to construct such a space, what its dimensions represent, or how many there might be. Here, to investigate such questions, the utility of E-space as a conceptual tool is explored further, with attention to the problem of representing diverse qualia in spaces of more than two dimensions.

This is not intended as a rigorous topological exercise, nor it seems, can it be, for reasons discussed below. Instead it is at this stage simply an investigation of a particular graphical construct as a tool for dealing conceptually with how phenomenal consciousness would unfold over evolutionary time in response to changes at the level of the SCs. Based on the ideas of von Békésy (1959), one can draw provisional conclusions regarding the nature of at least one E-space dimension: that among the properties to which mechanosensory qualia map (here combining tactile and acoustic experience), one of these properties will be time-related. More importantly, the analysis provides insights into the nature of ancestral experience prior to the emergence of a more differentiated form of consciousness, making the case that if evolution is to assign qualia to sensory modalities in an optimal way, the best starting point is to have ur-qualia that are diffuse and extend through many dimensions. A sorting process will then follow whereby different categories of qualia are progressively restricted to non-overlapping domains (i.e., exclusive sets of dimensions) in E-space. This provides insight into otherwise problematic issues, including how consciousness might have been experienced at different stages in its early evolution.

Exploring experience-space: Dimensionality and time

E-space (**Figure 1**) is designed to be an experiential counterpart to my configuration space representation of SCs. It was conceived as a way to map the qualitative properties of phenomenal experience so as to reflect the logic of how SCs influence that experience, so that axes in E-space would correspond to some combination of the neural features and/or events that characterize SC space. In this sense, there is no implied dualism, that E-space in some way represents a virtual realm separate from the material world. It is, instead, simply a mapping. And, as with SC-space, it applies only to qualia, conceived of as fundamental units of experience, and hence is unsuited for representing more complex contents that depend on sequential processing at a neurocircuitry level. So, for example, light perception can in principle be investigated using E-space, but not the visual display as a total experience.

E-space is constructed also to be ontologically fixed, in that it maps all qualia that could potentially exist in consequence of neural activity, whether experienced by any particular brain or

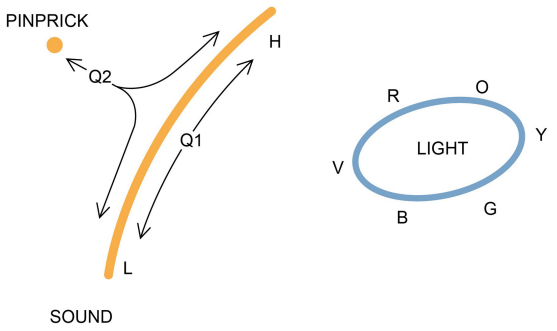


FIGURE 1
An experience space (E-space) representing three kinds of qualia in two dimensions, modified from Lacalli (2021). A pinprick, among the simplest of tactile experiences, and disregarding its localization, would be a point. Sound, for animals that can consciously distinguish pitch, would be a set of related qualia ranging from low (L) to high (H) pitch. Color, as we experience it, would be a closed curve, as the sequence from red to orange, yellow, green, blue, and violet (R, O, Y, G, B, V) is recursive, leaving the center of the curve for their blended combination, white light. The trajectories (arrows) show possible evolutionary sequences: that the pitch range of acoustic experience could originally have been limited to an acoustic ur-quale at Q1, and then have expanded over time; or that both tactile and acoustic sensations could have a common origin in an intermediate ur-quale (at Q2) that combined features of both, making the descendant qualia homologous as experiences (see Lacalli, 2022 for further comment on homology at the experiential level). The SCs responsible for evoking intermediate sensations along the trajectories will have been extinguished by selection, but will have existed in the past whereas, for qualia unrelated through homology, there may be no such intervening points, and hence no access to intermediate experiences. This may be the case for light and mechanosensations, which share no obvious qualitative features, in which case there would be no justification for mapping them to the same surface.

not. As such, it represents a fixed domain of possibilities that evolution explores through neural innovation, encountering qualia of adaptive utility in much the same way that an exploration of the various mineral elements available in sea water would identify calcium as the one most suitable for constructing shells and skeletons. E-space therefore differs from topological constructs used to map empirical data on conscious experience based on subjective reporting, including similarity space (Raffman, 2015) and quality space (Rosenthal, 2015), which are, in any case, not designed to address the problem of evolutionary change. And, though subjective reporting is used here as a guide to constructing the figures, e.g., in the choice of acoustic pitch and visual hue as variables for mapping, this choice is provisional, and may require revision once data are available on real SCs, as opposed to hypothetical ones, and the way they map to experience. The main conclusions of my analysis are, in any case, of a general nature, and valid irrespective of the specific details of how E-space axes are defined.

Two mechanosensory modalities are included in **Figure 1**, a pinprick, to represent sharp pain, and sound, along with the perceived spectrum of light. These are chosen to provide two separate demonstrations of why E-space must have at least two dimensions. For mechanosensations, this is because deriving both tactile and acoustic qualia from a common ancestral ur-quale requires divergence along two trajectories, which then define separate axes in E-space, one of which can be provisionally assigned to represent pitch. Assuming the range of perceived pitch has expanded over evolutionary time, a trajectory would then be traced out approximating that shown in the figure, beginning at the point (Q1) representing the ancestral acoustic ur-quale. A step further is to suppose a degree of homology between acoustic and non-acoustic mechanosensations, and derive both from an ur-quale (Q2) intermediate between them that combines features of both. This yields a second, independent axis, and the points traced out by this divergence then define a surface of two dimensions at a minimum that, assuming evolutionary change is incremental, is locally continuous. Intermediate points along such trajectories can then be considered real, i.e., they exist, because they have existed in the past in real brains.

A digression is required here on terminology, as to what points in E-space represent. Since E-space is designed to map qualia conceived of as fundamental units of experience, there is a potential problem in supposing they can be assigned subsidiary properties like pitch. The solution to this problem is to treat each point in E-space as representing a single quale, and the subsidiary "properties" as labels that define the relation between a given quale and its close neighbors. The curves and lines in the figure representing modes of sensory experience (sound, light, etc.) are then sets of related qualia, and the domains they occupy (the points, lines and curves in the figure, which can be diffuse or compact) are point clouds that map these sets of qualia. However, to be consistent with previous usage (in Lacalli, 2021), I will use the singular "ur-quale" to refer to ancestral ur-experiences conceived of as point clouds that may combine in one experience the properties of what we would identify as belonging to distinguishable qualia.

Light perception is shown in **Figure 1** as two-dimensional because mapping the recursive feature of light experience, where hues blend into each other to form a color wheel, also requires a minimum of two dimensions. This is adjusted in **Figure 2**, so the two defining axes are yellow/blue and red/green in accord with current theory (Matthen, 2020), but the question of more immediate concern is whether it is appropriate to represent light qualia on the same surface as mechanosensations. The alternative is for E-space to be multidimensional where n, the number of dimensions, is large (i.e., it is a large dimensional space, potentially with at least several times as many dimensions as there are categories of qualia to occupy them), in which case we have a formal construction that is more difficult to illustrate but far richer in what it can be used to represent. And, since

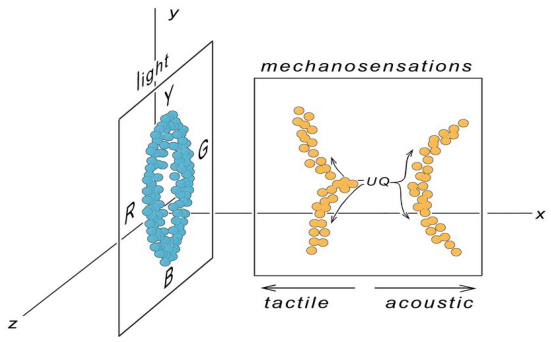


FIGURE 2
The E-space from **Figure 1** reconfigured for 3 dimensions, with mechanosensations and light qualia each occupying their own 2D surface oriented in one of many possible ways. Experiences are shown as point clouds, less compact than in the previous figure, as a reminder that each point in E-space represents the action of a selector circuit (SC) responding to a particular sensory stimulus. The overall density of points is a reflection of how many SCs are dedicated to each sensory modality, and would undoubtedly, for a fully evolved consciousness, be much greater than shown. Here, strictly as a thought experiment, acoustic and tactile sensations are supposed to originate from a common ur-quale (UQ), and share a common axis representing pitch (the *y* axis) in accord with the ideas of von Békésy. In addition, the experience of pain (the pinprick in **Figure 1**) has been expanded so as to encompass both sharp and dull pain along the same axis as acoustic pitch. Orienting the light plane along this same axis would imply that the yellow/blue axis is pitch-like in some way, while the red/green axis is not. But trajectories along the tactile/acoustic axis intersect the plane for light experience, implying a link between these. As discussed more fully in the text, such implied relationships are difficult to justify, yet are an inescapable feature of mapping more than one category of experience to a space with too few dimensions. Allowing E-space to have more dimensions, perhaps many more dimensions, avoids this problem.

the dimensions are simply hypothetical axes along which the separable properties of experience are mapped, in other words the degrees of freedom for the system, there is no reason *a priori* to limit to their number. This also means dimensions in E-space will differ from those of normal 3D space in that continuity across them is not guaranteed so that, with reference to **Figure 1**, there may be no route by which a light experience can transition incrementally into a mechanosensation. Assigning qualia to separate sets of dimensions avoids this problem, but there could still be discontinuities between dimensions, which is an impediment to investigating E-space as a whole using mathematical tools requiring continuity.

Before considering arbitrarily large dimensional spaces, a further digression is useful on the problems that arise from having too few dimensions. This is illustrated in **Figure 2**, which expands the first figure from two to three dimensions. Mechanosensations and light qualia are now represented as restricted to separate planes, with the pinprick-related (tactile) trajectory extended along the *y* axis, so that sharp and dull tactile experiences diverge from a common origin along an axis parallel to that for acoustic pitch. This accords with the classical proposal by von Békésy (1959, 1960; see also Tonndorf, 1986; Manley et al., 2012) that longer wavelength components in the stimulus, whether for sound or mechanosensations more generally, correspond to sensations that are lower pitched and spatially less focused. The *y* axis in the figure would then be a measure of something related to wavelength and frequency, i.e., time, which would not mean time itself, as in the duration of the experience, but some other time-dependent feature encoded in neural activity. Von Békésy's proposal is useful for illustrating the point that axes in E-space are most easily understood when we have at least a provisional idea of the neural basis for positional shifts along those axes. Yet in most cases this will not be even remotely the case, as to the neural basis of the difference between the sensation of red and green, or yellow and blue, for example, and whether differences along the axes defined by those hues depend on related neurocircuitry features or not.

Consider now what happens if we try to use the pitch axis (the *y* axis in **Figure 2**) for another set of qualia, namely light perception. The two planes, for mechanosensations and light, could in principle be oriented in various ways in a three-dimensional space, but the point is made by examining two cases, where the planes are either perpendicular or parallel to one another. Take first the perpendicular case, shown in the figure, with mechanosensations and light mapped to planes aligned along the *xy* and *yz* axes, respectively. This implies that both share similar properties across the *y* axis, but otherwise not. As drawn, the shared axis relates acoustic pitch to the yellow/blue axis for light, which would imply that there is something intrinsically "higher pitched" about yellow as compared with blue, but also that this same property could not be used to distinguish red from green. This privileges one set of hues over another, as being more sound-like, which begs the question of how likely it is that distinctions applicable to one sensory modality (here, high vs. low pitch) will apply to others. There is first the problem of separating the quality of an experience from its intensity, for example, in the case of affect (see Cabanac, 2002), whether a strongly felt emotion is one that is more narrowly focused in a pitch-like sense, or simply more intense. At the level of SCs, differences in intensity might simply be a matter of circuit redundancy, with intensity increasing in proportion to the number of SCs available for activation. But consider hedonicity, another of Cabanac's properties: does that define an axis shared between non-homologous contents, so that a pleasant odor and a sense of contentment might be supposed to depend on a common mechanism at the level of SCs, or not? Though the idea that it does may have some appeal, comparisons of this kind between non-homologous contents have an intuitive component conditioned by the language we use to describe experience that may be quite misleading (Walker et al., 2012; Van Leeuwen et al., 2015), which for my analysis makes the issue as a whole sufficiently problematic that it is better deferred. So, returning to the figure, note the further problem that trajectories

along the tactile/acoustic axis for mechanosensations (the *x* axis) intersect the plane representing light experience, implying that whatever separates tactile experience from sound, the more you have of it in one direction or the other (depending on whether the mechanosensory plane is rotated around the *y* axis, or not), the closer you get to a light experience. Absolute distances in E-space are not specified, and very large distances could conceivably account for apparently dissimilar experiences being related in this way through a shared dimension, but it is still a stretch to suppose that a transition through incremental steps is possible between experiences as different as sound and light.

The case of parallel planes can be visualized by rotating the plane for mechanosensations in **Figure 2** by 90 degrees along the *y* axis, so it parallels the light (*yz*) plane, but at a different values *x*. The time-related *y* axis is still shared, but now, along the *z* axis, differences between tactile and acoustic experience and red versus green hues would depend on the same property, meaning pain would differ from sound in the same way red differs from green. This is rather puzzling, because differences between yellow and blue are still shown as being frequency dependent, i.e., pitch-like, whereas there is no obvious difference between the experience of yellow vs. blue compared with red vs. green to suggest they differ fundamentally in this way. In sum, the mental gymnastics required to fit diverse sets of qualia into a small dimensional space raises more questions than it answers. The alternative, a more fruitful approach in my view, is to assume E-space extends across many more than three dimensions, and further, that few if any of these dimensions are shared between different categories of qualia as we experience them, as components of a fully evolved consciousness. How this situation would have evolved is a separate question, explored in the next section using the perception of light as an example, to argue for the operation of an exclusionary principle that facilitates both the divergence of qualia and their optimization for particular functions.

Large dimensional spaces: Light perception and the case for dimensional sorting

Light experience recommends itself to dimensional analysis because its recursive property cannot be represented in less than two dimensions. There is a long history of speculation on color perception, dating to Newton, but current thinking (Raffman, 2015; Matthen, 2020) explains the range of unique hues we experience as arising from the interactions between two principal color axes, yellow/blue and red/green, with a third for white vs. black. The subtleties of how hues are distinguished today is not, however, especially relevant to the evolutionary question of how this mode of color perception originated, because what then matters in biological terms is the ability to consciously distinguish light from the absence of light and from other forms of experience. And, while it is a valid evolutionary question to enquire whether conscious perception of light preceded the evolution of the ability to discriminate colors at the photoreceptor level, or the reverse, it does not matter when considering the first experience of light unless the ability to consciously perceive a full spectrum of hues was part of that first experience. Otherwise the perception of distinct hues would have been assembled later and incrementally, as the set of qualia we perceive as light was refined to implement that function. E-space can then be used as a framework for thinking both about this refinement process and about how light came to be perceived differently from other sensory modalities in the first place.

To this end, consider first an animal for which the perception of light has just emerged at a conscious level. This means at a material level that SCs capable of evoking a light experience are present. But what hue will they evoke, or, in other words, what are the characteristics of the ur-quale in terms of hue? The answer will depend on the redundancy of the system, meaning the number of active SCs required per brain to evoke a light experience. If one, then only one hue can be evoked at any one time, and this will vary between individuals in the population unless there is precise control at the SC level to ensure that each individual has replicated the same SC. But we would then need to account for why so precise a mechanism for specifying hue was already in place. Otherwise, with a less precise mode of specification, and hence a greater range of SCs at the population level, each individual would experience a different hue. Subsequently, assuming some hues or combinations of hue are better adapted for vision than others, selection would ensure those hues or combinations of hues became the population standard. More likely is a degree of redundancy, of multiple light-evoked SCs per brain, so the ur-quale for light for each individual would combine the experience of various hues, but in different ways (the point clouds would differ between individuals) so that individuals would have a similar but not identical experience. But there is then a further problem, assuming a degree of redundancy, as to whether the ur-quale for light would have been restricted to light-like sensations alone. It could instead have extended as a point cloud into regions of E-space supporting experiences that for us are associated with other sensory modalities, resulting in a mixed experience incorporating features we would recognize as belonging to those other modalities. The ur-quale for light would then differ from the pure experience of light as we perceive it, but would still have adaptive utility so long as it represented an improvement on the way light was perceived up to that point. This is because the well-known aphorism relating to vision, that "in the land of the blind, the one-eyed man is king" applies at every step in the evolutionary sequence, which is a further reminder of how distant our own

consciousness today may be from subjective experience as it first emerged in evolution.

Problems like those just mentioned are simplified if we think more clearly about how an emerging ur-quale would be represented in a large dimensional space. At the level of SCs, selection will act to increase the reliability with which a given quale is evoked so as to better distinguish it from other forms of emerging experience. This means point clouds in SC-space will become more compact with time (Lacalli, 2021), which for E-space, translates into a reduction in the dispersion of point clouds across dimensions. So the end point for the evolution of light perception, at least for us, would be its restriction to just a few dimensions, namely the ones we identify as light-like based on our own experience. At issue is the starting point, of whether the ur-quale for light was initially widely dispersed across E-space dimensions or restricted to just a few. This is equivalent to asking whether, with reference to SC-space, we are dealing with a "puddle" scenario described in the paper just cited (see figure 3 in Lacalli, 2021), of an ur-quale that combines multiple forms of experience that later came to be experienced separately, or the "tree" scenario, where qualia are precisely specified from the start. Again, redundancy matters because, when it is low, individual experience would differ due to few SC- and E-space points per brain being scattered in diverse ways across the dimensions occupied by the denser point clouds mapping that same ur-quale for the population as a whole. But so long as there is some redundancy at the individual level, meaning multiple SCs per individual, the starting point for the E-space counterpart of the puddle scenario at both the individual and population level can be thought of as a diffuse point cloud with components resident in many E-space dimensions. The set of qualia we associate with a particular sensory modality, light perception in this example, would not be accidentally "discovered" by evolution, but would have been present as a sub-component in the ur-experience from the start. Evolution can then extract that subcomponent by systematically removing from the population those gene variants responsible for the SCs evoking E-space points in dimensions other than those that are light-like. The tree scenario poses more of problem, because an explanation is then required for why a particular ur-experience would already have been so precisely specified as to be restricted to few dimensions *before* selection had an opportunity to act on it as a manifestation of an emergent consciousness.

The argument is most easily appreciated by consulting **Figures 3**, **4**, which are designed to deal with the most general case, of qualia evolving simultaneously, and of emergent SCs on which selection has only just begun to act. The SCs can then be supposed not to be as precisely specified as they eventually will be, as a consequence of selection, which translates in E-space into point clouds that are more diffuse and spread across more dimensions than they eventually will be. **Figure 3** shows three coordinate axes that I will designate as representing light-like properties, though initially, as pointed out above, we could be dealing with a situation where light stimuli evoke points in other dimensions as well, perhaps many other dimensions. The figure then follows the conversion of an initially diffuse point cloud (**Figure 3A**, in blue), representing the ur-experience of light perception in the dimensions shown, evolving (in **Figure 3B**) into a flattened disk centered on the point in E-space corresponding to white light, defined as the point where all other hues are extinguished. At the same time, any other ur-experiences incorporating light-like properties will find those properties progressively eliminated. Hence, the red and orange dots in the figure, representing points in the light-like dimensions of E-space evoked by SCs in response to olfactory and acoustic stimuli, respectively, have either vanished from those dimensions at a later stage in evolution (orange dots in **Figure 3**), or are in the process of doing so (red dots). **Figure 3** provides no indication of what hypothetically might be happening in other E-space dimensions, but **Figure 4** does, for three other dimensions chosen from among those mapping odor-like properties. In this case, over a time interval comparable to that in **Figure 3**, it would be the odor-like properties of acoustic and light ur-experience that are progressively eliminated as olfactory experience is refined. Selection would thus be acting simultaneously in this scenario to extinguish the maladaptive light-like features from non-light experiences, and maladaptive odor-like features from non-olfactory experiences.

The point of the above line of argument is not that a justification is needed for the adaptive properties of the set of qualia employed for the perception of light, or any other sensory modality, but that there is a particular way for evolution to extract and refine those properties that allows for the most suitable set of qualia for each modality to be selected over all others. For light in particular, this would also account for how the experience of white light became the default for a combination of other hues: that if there is such a point in E-space, where all other hues are extinguished and replaced by a single hue to which they all converge, then that is where evolution will choose to center the point cloud representing light experience. The analysis does not purport to explain why there should be such a point, but so long as it exists, it explains how evolution comes to select that point over all others. It may be, however, that any closed loop in E-space generates such a point, so qualia other than light-like ones could be used for representing the light spectrum in a recursive fashion. And if so, we are no wiser than before as to whether our experience of light is uniquely suited to this purpose or not.

An objection to the above line of argument is that having emerging qualia share dimensions, and hence properties, is unrealistic if qualia are only useful as contents if they are clearly distinguishable from one another from the start, as they are today in our own consciousness. This implies precisely specified domains in SC-space as consciousness first evolved, which equates to the tree scenario referred to above, and to

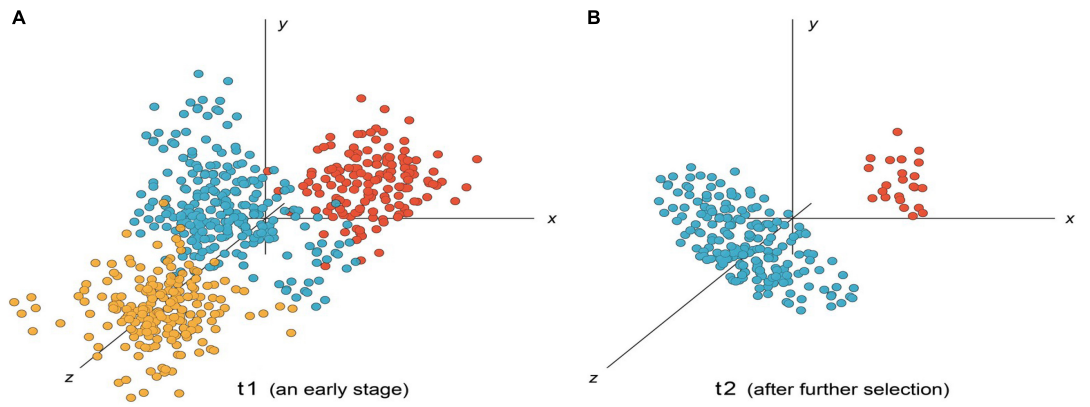


FIGURE 3
Evolving point clouds in a large dimensional version of E-space, showing three of those dimensions (*x*, *y* and *z*) that, for the sake of argument, are assumed here to map light-like experiential properties. There are of course many other dimensions that cannot be shown, and the ur-experience of light could well evoke points in those other dimensions, just as point clouds for other modalities might initially intrude, as shown here, into light-like dimensions. **(A)** Shows an early stage in the evolutionary process for a species for which consciousness is newly emergent from the preconscious condition, and **(B)** a later stage, after natural selection has had an opportunity to further refine that emergent set of experiences. The point clouds evoked in response to light stimuli are shown in blue, and those for two other modalities represented in *xyz* space in this hypothetical example, sound and odor, in orange and red, respectively. Through selection, evolution will progressively eliminate SCs evoking experiences that are maladaptive for each sensory modality, so in **(B)** point clouds evoked by stimuli other than light will have either been eliminated from these three light-like dimensions, as in the case of acoustic experience (the orange point cloud has vanished), or will be in the process of being eliminated, as here for olfactory experience (the now much smaller red point cloud). The point cloud for light experience, meanwhile, has become more compact so as to form a disc with white light at the center, as in the previous figures.

restricted dimensionality in E-space. But an explanation is then needed for how the SCs came initially to be so precisely specified. A possible answer, if we consider a single category of qualia evolving in isolation, is that there could be subsets of dimensions so superior in adaptive terms compared with the alternatives, that the restriction of an initially diffuse point cloud to those few dimensions occurred so rapidly as to be indistinguishable from its being precisely specified from the start. This would have consequences, especially where there is a sequence in which contents are added to consciousness as it evolves. To take a specific example, suppose light was the first sensory modality to be experienced consciously, in accord with the scenario suggested by Feinberg and Mallatt (2016). With light there is the added problem of whether we are dealing only with an experience that distinguishes consciously between light and the absence of light, or whether some form of consciously perceived 2D visual display was there from the start. But regardless, the relevant issue from an E-space perspective is that, once the point cloud evoked by light has been reduced to a suitably small number of light-like dimensions, contents added later to consciousness would evolve in a setting in which those few dimensions were already committed (i.e., occupied) and unavailable to any modality other than light perception. Hence SCs evoking a light-like experience for stimuli other than light would be strongly selected against and insignificant at a population level. The situation would be one of contents being added to consciousness in sequence, each in turn staking out a small subset of whatever uncommitted dimensions remain.

There are, however, other scenarios to be considered, including ones where there might initially have been no great advantage to selecting one set of dimensions over another for a given modality, or even to distinguish between modalities. So, for example, consider a rudimentary conscious arousal mechanism based on light and odor signals that used a nearly identical set of qualia for both. Assuming both were equally relevant signals for the initiation of a consciously controlled avoidance behavior, assigning them the same or a very similar quale would be perfectly adaptive compared with having no conscious input into the avoidance response from either modality. Differentiating the two modalities (light and odor in this example) might occur quite rapidly if there was an adaptive advantage to doing so, but there could otherwise have been a prolonged period when both were experienced in essentially the same way.

For my purposes in this account, the relative merits of any one such scenario or set of initial conditions over others is of less concern than ensuring that the broader framework, of mappings to E-space, is applicable to as wide a range of scenarios and initial conditions as possible. This would include scenarios where emotional feelings (positive and negative affect) are crucial to the narrative (e.g., Damasio and Carvalho, 2013; Solms, 2019), and modalities associated with the organs of

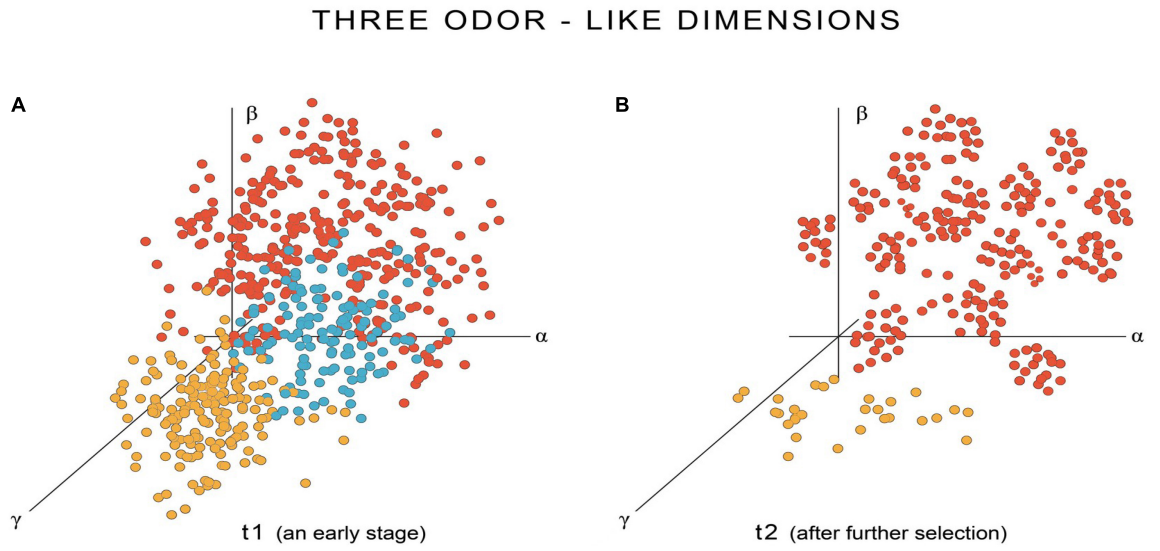


FIGURE 4
As in Figure 3, but a second set of point clouds in three other E-space dimensions (α, β, and γ) that map odor- rather than light-like properties. Blue, orange and red point clouds represent the odor-like components of ur-qualia for light, sound, and odor, respectively, as before. In **(A)** again an early stage in the evolutionary process of refinement, there is overlap between the three modalities, so that experience of both light (blue) and sound (orange) will have some odor-like properties. In **(B)** a later stage, the point cloud for odor has evolved into a clustered array, whereas odor-like features of light experience have been eliminated, and for acoustic experience, nearly so. The result overall is that the properties of qualia are sorted so as to optimize the way each is experienced while, at the same time, eliminating overlap between different categories of experience. The clustering of the olfactory point cloud in **(B)** unlike the disc used for light in the previous figure, is meant to indicate that, for olfaction, arranging for the sensations associated with related olfactory experience to be clustered, and perhaps to extend to many more dimensions, might be a way to enhance the ability to store and recall related odor experiences from memory in a systematic way. The point is made simply as a reminder that the shape and dimensionality of point clouds in E-space will likely differ, perhaps dramatically so, between different sensory modalities.

special sense in consequence get correspondingly less attention. With the above caveats in mind, and deferring the complications inherent in special cases, I feel justified in concluding this section with the following conjecture for the general case of ur-experiences evolving together: that if it can be assumed that qualia are assigned to sensory modalities in ways that are either optimal or better than the alternatives, the most effective means of achieving this in a systematic fashion is for the ur-qualia for these modalities to begin the process as diffuse, multidimensional point clouds in E-space. This provides evolution with the widest range of options, and so avoids the problem of assigning a less-than-optimal quale by default, simply because that happened to be the way a given modality was first experienced, or because all other dimensions were already committed to other modalities. And because, for the general case, diffuse, multidimensional ur-experiences offer this advantage over narrowly specified ones, one can predict that taxa whose brains employ the diffuse option are the ones that are most likely to have survived to the present. Hence, the qualia their brains experience are more likely than not to have been selected in this fashion. For the selection process as a whole, I suggest the term "dimensional sorting" to emphasize this outcome: that an optimal sorting of qualia among available dimensions can, by this means, be achieved. In addition, and very importantly, if we can assume the sorting process occurs gradually over time, and impacts most if not all emerging contents simultaneously, this model for the process can account for the evolution of conscious experience as a balanced, unified whole. This is because contents evolving together as an ensemble are continuously being tested for their effect on the totality of experience as the sorting process proceeds.

The experience-space/selector circuit-space relationship, and the exclusionary principle

The above analysis makes the case that divergence and optimality among qualia are facilitated by having ur-qualia occupying many E-space dimensions so that multiple distinguishable properties can be sorted among qualia. There is also then an exclusionary principle in operation, that evolution will act to prevent qualia from incorporating properties evoked by other qualia. The exclusionary principle applies in this case across all available dimensions, and should serve in practice to distribute qualia as widely as possible across those dimensions.

Divergence and the exclusionary principle also operate in SC-space, but there they act within dimensions, to maximize

distance and minimize overlap between point clouds on a dimension-by-dimension basis. This distinction is worth bearing in mind when dealing with mappings from SC-space to E-space. There are cases where an isomorphic mapping is possible, for example, for closely related (i.e., homologous) qualia such as the experience of different acoustic tones (Lacalli, 2021). Minor adjustments to the SCs might in that case be sufficient to generate meaningful change within the E-space dimensions that define acoustic experience, so the mapping would be from one low-dimensional space to another. However, for change involving non-homologous forms of experience, such as a transition from an acoustic experience to one that is light-like, an isomorphic mapping seems the least likely alternative. This is because selection acts on point clouds in SC-space so as to maximize configurational differences within dimensions, but there is no corresponding benefit to reducing dimensionality *per se*. In contrast, the result in E-space will be seen predominantly in the restriction of point clouds for each category of experience to a small subset of dimensions, so the shapes of point clouds across dimensions in E-space are being changed in a fundamentally different way than in SC-space.

To go further with the evolutionary argument, there are plausible conclusions to be drawn, given suitable assumptions, as to how subjective experience would have changed as consciousness first evolved. Here I take the simplest case, of an explicitly neurophysical stance: that the evolutionary precursor of subjective experience arose from some physical consequence of neural circuit activity, which equates to "the physical" (Godfrey-Smith, 2019; Jylkka and Railo, 2019), or a neuroscientific point of view (Winters, 2021). This, in some formulations, is attributed to underappreciated properties of electromagnetic fields (McFadden, 2020; Kitchener and Hales, 2022), but regardless of details, the point is that a neurophysical stance gives meaning to the idea of redundancy, that it involves replicate circuits acting in concert. If sentience then depends on circuits exhibiting a degree of redundancy, the expectation is that those circuits would have been neither numerous nor very effective in producing sentient experience until evolution was able to further augment that experience and refine it. In other words, the initial rudiment of phenomenal experience present in the emerging conscious state would, for the individual, have been of low intensity and comparatively undifferentiated. The action of evolution would then have been twofold: to increase the intensity of the experience while, at the same time, beginning the dimensional sorting process, of extracting subcomponents and increasing their intensity individually. This would presumably have depended on increasing the redundancy of the system as a whole, because that is the only way of augmenting the raw material, at the circuitry level, on which selection acts. For the individual, there should therefore have been an increase in the intensity of experience over time from an initially negligible level, but also a transition from an undifferentiated noise-like form of experience, to one where one or more distinguishable contents emerged from this noisy background. And, for species for which consciousness is newly emergent, assuming this primarily involves qualia as opposed to more complex contents, the conscious state would be something evoked by specific stimuli, and so would have been more episodic than our own, whose complex formatted contents (e.g., vision and abstract thought) are adaptive in large part because they occupy the mind, when awake, on a more-or-less continuous basis.

Conclusions, with caveats

This account is concerned with the evolution of consciousness, and while there are various ways of addressing the issue (e.g., Cabanac et al., 2009; Velmans, 2012; Feinberg and Mallatt, 2016; Gutfreund, 2018; Ginsburg and Jablonka, 2019; Godfrey-Smith, 2019; Black, 2021; Lacalli, 2022), the focus here is on how selection would act on the simplest contents of consciousness as they first began to evolve. Though the hard problems of consciousness enter the narrative at various points, they are not addressed directly, my view on the subject being (in accord with Block, 2009), that physics may hold the answer, but we currently lack the data and conceptual tools needed to discover that answer. But in any case, the questions one can address in evolutionary terms are less concerned with how consciousness can exist than how it got to be the way it is, and the constraints that govern its evolutionary trajectory along particular paths as opposed to others. Current theories of consciousness are diverse in their focus and claims (e.g., Atkinson et al., 2000; Van Gulick, 2018; Seth and Bayne, 2022), but the role evolution plays in determining the character of phenomenal experience is seldom dealt with as explicitly as one would like, especially by higher order theories. Yet, if we take Dobzhansky's dictum with the seriousness it deserves, that nothing in biology makes sense except in the light of evolution (Dobzhansky, 1973), then dealing with evolutionary issues like sequence and homology is an essential part of understanding how an evolved consciousness such as ours came to be the way it is. The formulation presented here has one advantage in this respect, that it focusses attention on how the properties of experience, expressed in dimensional maps, will have changed over time, and hence on how the experiences of our distant ancestors might have differed from our own. This would include such arcane questions as to whether, for example, our species would, in its history, have had access to sensations comparable to those experienced by, say, an electric fish during an electric discharge, or a bat as it echolocates.

Certain caveats should be kept in mind with the E-space formulation as developed here. First, that it is an awkward fit for theories where qualia are not separable from consciousness as a unified whole, and hence are not individually subjects of selection (Brook and Raymond, 2021). But such theories present

difficulties to an evolutionary analysis of any kind, which leaves them largely outside the concerns of evolutionary biology, and hence of this account. But even for theories of consciousness where E-space would in principle be applicable, there is a question as to how useful it is for dealing with the realm of experience. One can ask, for example, whether E-space is well founded as an analytical tool. But this is difficult to assess until we have a better understanding of the nature of the properties being mapped in this exercise including whether, for example, axes in E-space are orthogonal, as spatial dimensions would be, or can be made so. Hence, without knowing precisely what E-space axes represent, there is no guarantee that E-space has the features required for mathematical analysis, of orthogonality and continuity, or whether it has any meaning beyond being a device for ordering empirical data in graphical form.

One can nevertheless argue, at a minimum, that E-space is worth exploring if it provides insights beyond those available from more conventional forms of narrative and verbal argument. From my analysis there appear to be two such insights. First, the question of shared axes highlights the importance of ideas like those of von Békésy, uniting sound with other mechanosensations, the implication being that at least one E-space axis must be time-related. A testable prediction would then be that there are common time-related features at the neurocircuitry level shared across mechanosensory SCs, which could be proved or disproved from sufficiently detailed data on brain circuitry once such data are available. Other axes in E-space are more problematic, e.g., for light, and I can in consequence offer no useful comments on, for example, how a yellow/blue or red/green axis relates to SC structure or activity patterns. There is a further problem of hidden structure in E-space, which can again be illustrated using light perception: that using two axes to represent the observable range of hues may simply mean that a point cloud occupying more dimensions than two is experienced as if it were projected onto a 2D surface. So in **Figure 3B**, for example, a flat disc is used to represent light experience, yet it resides in a larger dimensional space, of three dimensions in this example, though there could conceivably be more. Our perception of hue being defined by two axes, of yellow/blue and red/green, would then, in effect, be a matter of the brain making some form of secondary coordinate transformation.

The second insight, at a more abstract level, is valid irrespective of how E-space dimensions are defined in practice. It is that unrelated sets of qualia are best represented in E-space by mapping them to non-overlapping sets of dimensions and, flowing directly from this formulation, that the assignment of qualia to sensory modalities is most efficiently achieved for contents evolving together if the respective ur-qualia are initially diffuse and multidimensional. This expands the pool of options on which evolution can draw, and is not only the better strategy from the standpoint of adaptive flexibility, but provides the best available way of conceptualizing the process by which qualia are optimized by evolution for the functions they are required to perform. I refer here to the process of dimensional sorting as described above, whereby diffuse multidimensional ur-experiences will have an evolutionary advantage over those more narrowly specialized from the start. This also resolves a philosophical question (e.g., see Majeed, 2016), of how is it that a particular assortment of conscious contents can be brought into existence. The question is inescapably an evolutionary one, for which the answer is straightforward if one assumes the process begins with ur-experiences consisting of separable components, because all that remains is for evolution to effect the separation in ways that are functionally useful.

Data availability statement

There are no data beyond that included in this article. Further inquiries can be directed to the corresponding author.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Funding

This work was supported by the L. G. Harrison Research Trust.

Acknowledgments

The author thank the reviewers for their comments, and Riley Lacalli for preparing the figures.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Atkinson, A. P., Thomas, M. S. C., and Cleeremans, A. (2000). Consciousness: mapping the theoretical landscape. *Trends Cogn. Sci.* 4, 372–82. doi: 10.1016/S1364-6613(00)01533-3

Bedau, M. A. (1997). "Weak emergence," in *Philosophical Perspectives: Mind, Causation, and World*, Vol. 11, ed. J. Tomberlin (Malden, MA: Blackwell), 375–99.

Black, D. (2021). Analyzing the etiological functions of consciousness. *Phenom. Cogn. Sci.* 20, 191–216. doi: 10.1007/s11097-020-09693-z

Block, N. (2009). "Comparing the major theories of consciousness," in *The Cognitive Neurosciences IV*, Chap 77, ed. M. S. Gazzaniga (Cambridge, MA: MIT Press), 1111–23.

Brook, A., and Raymond, P. (2021). "The unity of consciousness," in *The Stanford Encyclopedia of Philosophy*, ed. E. N. Zalta (Stanford, CA: Stanford University).

Cabanac, M. (2002). What is emotion? *Behav. Processes* 60, 69–84. doi: 10.1016/S0376-6357(02)00078-5

Cabanac, M., Cabanac, A. J., and Parent, A. (2009). The emergence of consciousness in phylogeny. *Behav. Br. Res.* 198, 267–72. doi: 10.1016/j.bbr.2008.11.028

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *J. Cons. Stud.* 2, 200–19. doi: 10.1093/acprof:oso/9780195311105.003.0001

Damasio, A., and Carvalho, G. (2013). The nature of feelings: evolutionary and neurobiological origins. *Nat. Rev. Neurosci.* 14, 143–52. doi: 10.1038/nrn.3403

Dobzhansky, T. (1973). Nothing in biology makes sense except in the light of evolution. *Am. Biol. Teach.* 35, 125–9. doi: 10.2307/4444260

Feinberg, T. E., and Mallatt, J. M. (2016). *The Ancient Origins of Consciousness*. Cambridge, MA: MIT Press.

Ginsburg, S., and Jablonka, E. (2019). *The Evolution of the Sensitive Soul: Learning and the Origins of Consciousness*. Cambridge, MA: MIT Press.

Godfrey-Smith, P. (2019). Evolving across the explanatory gap. *Philos. Theor. Pract. Biol.* 11:1. doi: 10.3998/ptpbio.16039257.0011.001

Gutfreund, Y. (2018). The mind-evolution problem: the difficulty of fitting consciousness in an evolutionary framework. *Front. Psychol.* 9:1537. doi: 10.3389/fpsyg.2018.01537

Hohwy, J., and Bayne, T. (2015). "The neural correlates of consciousness: causes, confounds and constituents," in *The Constitution of Phenomenal Consciousness: Towards a Science and Theory*, ed. S. M. Miller (Amsterdam, NL: John Benjamins Publ. Co), 155–76. doi: 10.1075/aicr.92.06hoh

Jylkka, J., and Railo, H. (2019). Consciousness as a concrete physical phenomenon. *Cons. Cogn.* 74:102779. doi: 10.1016/j.concog.2019.102779

Kitchener, P. D., and Hales, C. G. (2022). What neuroscientists think, and don't think, about consciousness. *Front. Hum. Neurosci.* 16:767612. doi: 10.3389/fnhum.2022.767612

Klein, C., Hohwy, J., and Bayne, T. (2020). Explanation in the science of consciousness: from neural correlates of consciousness (NCCs) to difference makers of consciousness (DMCs). *Phil. Mind Sci.* 1:4. doi: 10.33735/phimisci.2020.II.60

Lacalli, T. C. (2020). Evolving consciousness: insights from Turing, and the shaping of experience. *Front. Behav. Neurosci.* 14:598561. doi: 10.3389/fnbeh.2020.59856

Lacalli, T. C. (2021). Consciousness as a product of evolution: contents, selector circuits, and trajectories in experience space. *Front. Syst. Neurosci.* 15:697129. doi: 10.3389/fnsys.2021.697129

Lacalli, T. C. (2022). An evolutionary perspective on chordate brain organization and function: insights from amphioxus, and the problem of sentience. *Phil. Trans. R. Soc. B* 377:20200520. doi: 10.1098/rstb.2020.0520

Levine, J. (1983). Materialism and qualia: the explanatory gap. *Pac. Phil. Quart.* 64, 354–61. doi: 10.1111/j.1468-0014.1983.tb00201.x

Levine, J. (2009). "The explanatory gap," in *The Oxford Handbook of Philosophy of Mind*, eds A. Beckman, B. P. McLaughlin, and S. Walter (Oxford, UK: Oxford University Press).

Lycan, W. (2019). "Representational theories of consciousness," in *The Stanford Encyclopedia of Philosophy*, ed. E. N. Zalta (Stanford, CA: Stanford University).

Majeed, R. (2016). The hard problem and its explanatory targets. *Ratio* 29, 298–311. doi: 10.1111/rati.12103

Manley, G. A., Narins, P. M., and Fay, R. R. (2012). Experiments in comparative hearing: Georg von Békésy and beyond. *Hear. Res.* 293, 44–50. doi: 10.1016/j.heares.2012.04.013

Matthen, M. (2020). "Unique hues and colour experience," in *The Routledge Handbook of Philosophy of Colour*, eds D. H. Brown and F. Macpherson (London: Taylor & Francis), 159–74. doi: 10.4324/9781351048521-14

McFadden, J. (2020). Integrating information in the brain's EM field: the cemi field theory of consciousness. *Neurosci. Cons.* 2020:16. doi: 10.1093/nc/niaa016

Neisser, J. (2012). Neural correlates of consciousness reconsidered. *Cons. Cogn.* 21, 681–90. doi: 10.1016/j.concog.2011.03.012

Raffman, D. (2015). "Similarity spaces," in *The Oxford Handbook of Philosophy of Perception*, ed. M. Matthen (Oxford, UK: Oxford University Press), 679–93. doi: 10.1093/oxfordhb/9780196600472.013.030

Rosenthal, D. (2015). "Quality spaces and sensory modalities," in *Phenomenal Qualities: Sense, Perception, and Consciousness*, eds P. Coates and S. Coleman (Oxford, UK: Oxford University Press), 33–73. doi: 10.1093/acprof:oso/9780198712718.003.0002

Seth, A. K., and Bayne, T. (2022). Theories of consciousness. *Nat. Rev. Neurosci.* 23, 439–52. doi: 10.1038/s41583-022-00587-4

Solms, M. (2019). The hard problem of consciousness and the free energy principle. *Front. Psychol.* 9:2714. doi: 10.3389/fpsyg.2018.02714

Tonndorf, J. (1986). Georg von Békésy and his work. *Hear. Res.* 22, 3–10. doi: 10.1016/0378-5955(86)90067-5

Van Gulick, R. (2018). "Consciousness," in *The Stanford Encyclopedia of Philosophy*, ed. E. N. Zalta (Stanford, CA: Stanford University).

Van Leeuwen, T. M., Singer, W., and Nikolic, D. (2015). The merit of synesthesia for consciousness research. *Front. Psychol.* 6:1850. doi: 10.3389/fpsyg.2015.01850

Velmans, M. (2012). The evolution of consciousness. *Contemp. Soc. Sci.* 7, 117–38. doi: 10.1080/21582041.2012.692099

von Békésy, G. (1959). Similarities between hearing and skin sensations. *Psych. Rev.* 66, 1–22. doi: 10.1037/h0046967

von Békésy, G. (1960). *Experiments in Hearing*. New York, NY: McGraw-Hill.

Walker, L., Walker, P., and Francis, B. (2012). A common scheme for cross-sensory correspondence across stimulus domains. *Perception* 41:1186. doi: 10.1068/p7149

Winters, J. J. (2021). The temporally-integrated causality landscape: reconciling neuroscientific theories with the phenomenology of consciousness. *Front. Hum. Neurosci.* 15:768459. doi: 10.3389/fnhum.2021.768459

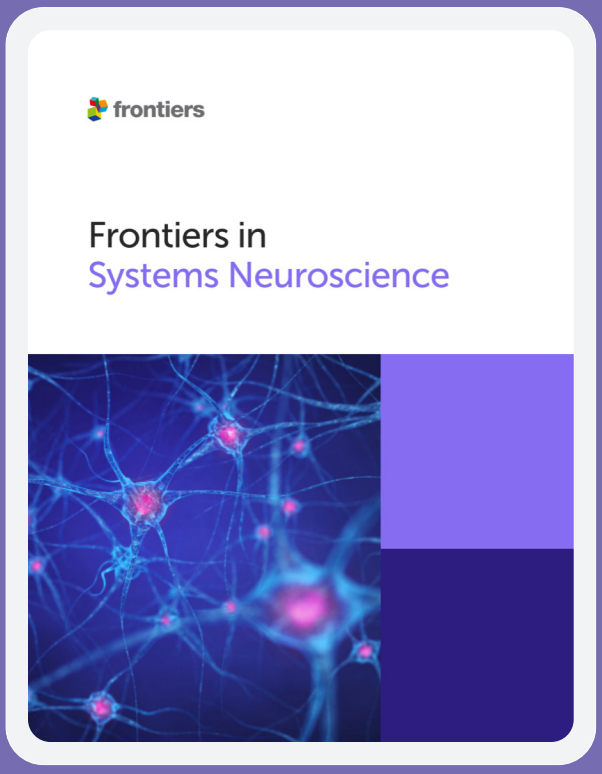
Frontiers in
Systems Neuroscience

Advances our understanding of whole systems of the brain

Part of the most cited neuroscience journal series, this journal explores the architecture of brain systems and information processing, storage and retrieval.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact