islands of awareness

Anil K Seth

Sackler Centre for Consciousness Science
University of Sussex
CIFAR Program on Brain, Mind, and Consciousness

NASEM Committee: Organoids and Chimeras

University of Sussex
Sackler Centre for Consciousness Science

defining consciousness

- consciousness is the presence of any kind of subjective experience whatsoever
- for a conscious organism, there is something it is like to be that organism
- consciousness ~= awareness ~= sentience

- consciousness is *not* coextensive with language, thought, intelligence, selfhood/identity
- consciousness is *not* coextensive with life, sensation, perception

Nagel (1974) *Philosophical Review*

defining consciousness

- conscious creature: is X capable of consciousness?
- conscious level: how conscious is X?
- conscious content: what is X conscious of?
- conscious self: subset of content - the experience of being a ‘self’

Seth (2016) *Aeon*

the extent of consciousness

- all (intact) humans are conscious (but not all of the time)
- other animals are conscious (but not all of them)
- artificial systems are (probably) not conscious (not yet)

- difficult cases:
    - non-mammalian animals (depends who you ask)
    - humans in coma, vegetative state etc.
    - organoids and chimeras
    - machines (not for now)

detecting consciousness via behaviour

- in humans, availability of *explicit* behavioural report
- in animals, utlise *implicit* behavioural report

Blindsight in monkeys

Alan Cowey* & Petra Stoerig†

* Department of Experimental Psychology, Oxford University, South Parks Road, Oxford OX1 3UD, UK
† Institute of Medical Psychology, Ludwig-Maximilians-University, Goethestrasse 31, 80336 Munich, Germany

BLINDSIGHT, the visually evoked voluntary responses of patients with striate cortical destruction that are demonstrated despite a phenomenal blindness, has attracted attention from neuroscientists and philosophers interested in problems of perceptual consciousness and its neuronal basis[1–3]. It is assumed to be mediated by the numerous extra-geniculostriate cortical retinofugal pathways whose properties are studied primarily in monkeys[4]. Like patients with blindsight[4–7], monkeys with lesions of the primary visual cortex can learn to detect, localize and distinguish between visual stimuli presented within their visual field defects[4,8–11]. Although the patients deny seeing the stimuli they can nevertheless respond to (by forced-choice guessing) in their phenomenally blind fields, it is not known whether the monkeys experience the same absence of phenomenal vision. To determine whether they too have blindsight, or whether they actually see the stimuli in their field defects, monkeys who showed excellent detection in tasks where a visual stimulus was presented on every trial, albeit at different positions, were tested in a signal-detection task[12] in which half the trials were blank trials, with no visual stimulus. They classified the visual stimuli presented in the field defect as blank trials, demonstrating, like patients, blindsight rather than degraded real vision.

If we want to understand why blindsight is blind, we need to know whether the monkeys, like patients, lose phenomenal vision as a result of striate cortical destruction. To answer this question, we studied four macaque monkeys. One, called Rosie, a female (*Macaca mulatta*) aged 8 years, had normal vision. The other three, Dracula, Lennox and Wrinkle, were older males (two *M. mulatta* and one *M. fascicularis*) who had had the striate cortex of the left cerebral hemisphere surgically removed and the splenium of the corpus callosum severed several years before the present experiment. All monkeys are still being studied behaviourally, but histological evidence from the excised tissue and a magnetic resonance scan showed that the ablation was complete (Fig. 1).

For three years their ability to detect and localize brief visual stimuli was studied while the monkey looked at the display shown in Fig. 2*b*. Eye movements were monitored (Fig. 2*a*), and a touchscreen recorded where the monkey touched the display. When a 2° white square of intensity 40 cd m^{-2} appeared at the bottom centre of the white screen (0.6 cd m^{-2}) and the monkey fixated and touched it, the square disappeared and was followed instantaneously by a brief (10–200 ms) 2° stimulus at one of the four positions 19° off the vertical midline. If the monkey touched the position at which the target had appeared, he/she was rewarded with a peanut or raisin. Two of these positions were within the normal left hemifield, the other two were in the right hemifield affected by the striate cortical lesion. Random responding would yield about 50% correct responses in either hemifield. By varying the intensity of the stimulus, we determined the increment thresholds for 75% correct performance. Compared with the normal side, they were elevated by 0.5 (Dracula and Lennox) to 1.5 log units (Wrinkle) in the affected hemifield. Note that a visual stimulus appeared on every trial, and that all monkeys performed at levels approaching 100% correct with stimuli 0.7 log units above threshold (Fig. 2*c*). When we conducted an entire session without presenting a visual stimulus (all blank

Cowey & Stoerig (1995) *Nature*

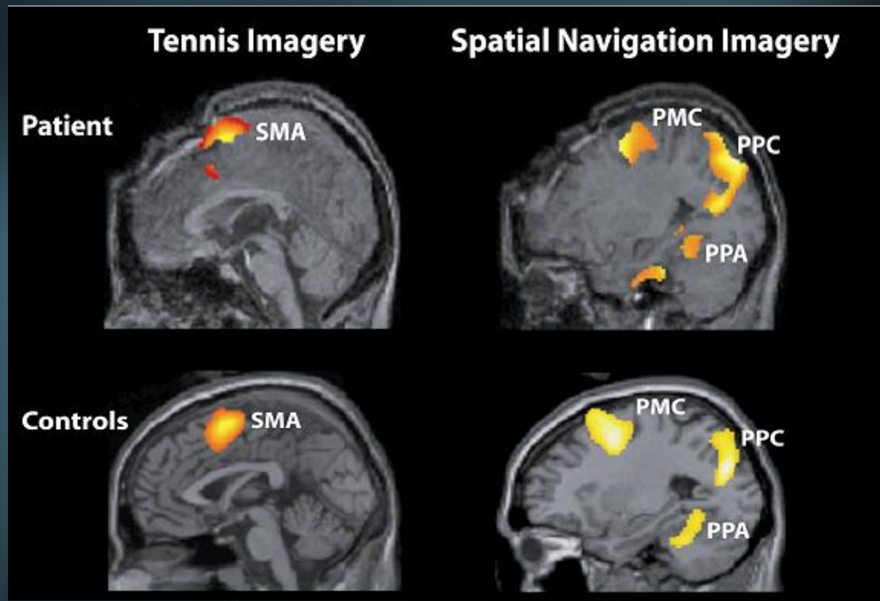
using humans as a benchmark

- the danger of anthropomorphism: mistakenly assigning conscious status on the basis of superficial similarity
- the danger of anthropocentrism: mistakenly denying conscious status on the basis of superficial dissimilarity



disorders of consciousness

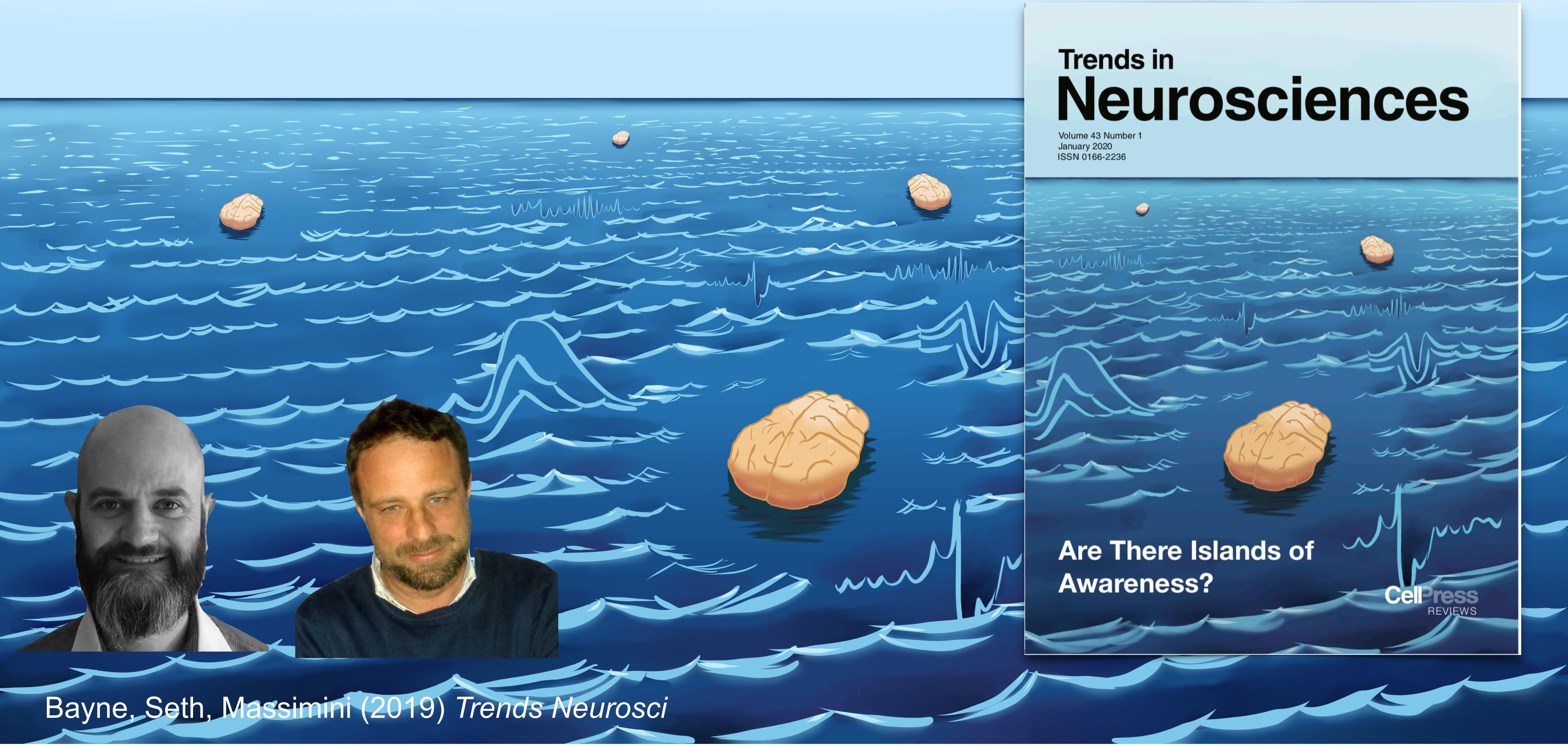
- conditions like the persistent vegetative state can lead to both false positives and false negatives



Owen et al (2006) *Science*

but what if there is no motor output or sensory input at all?

are there islands of awareness?



Bayne, Seth, Massimini (2019) *Trends Neurosci*

are there islands of awareness?

- is it possible for consciousness to occur when a brain (or parts thereof) is fully isolated from its environment (including the body)?

Bayne, Seth, Massimini (2020) *Trends Neurosci*

cerebral organoids

cerebral organoids

- things you can find in cerebral organoids:
- deep and superficial cortical layer neurons; forebrain, midbrain, and hypothalamic regions; functional nerve tracts; mature oligodendrocytes; mature excitatory and inhibitory neurons, etc

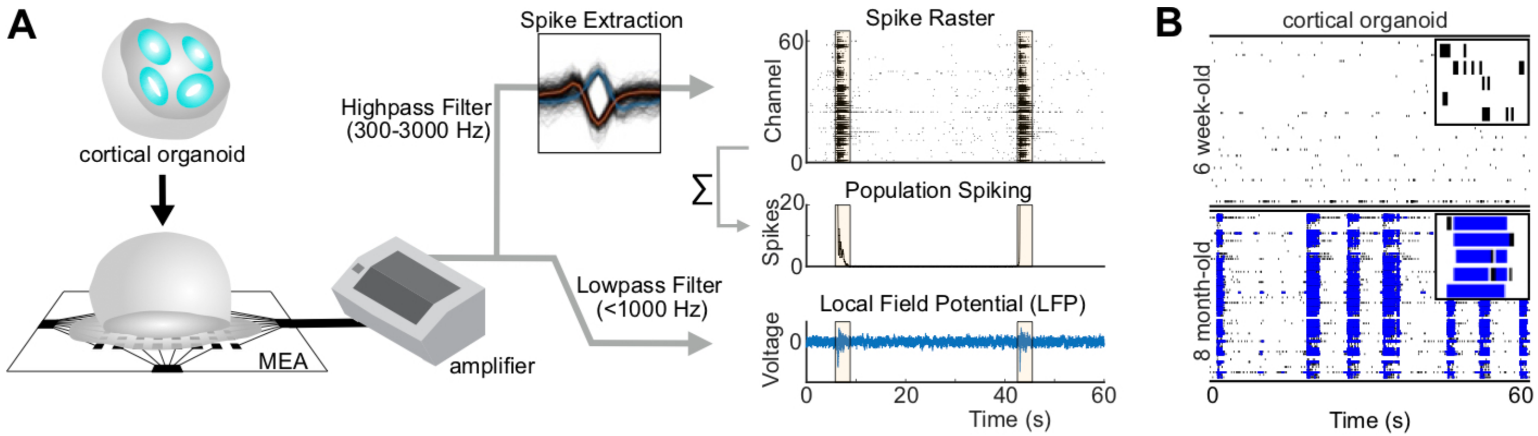
- things you don't (yet) find:
- well-defined neuroanatomical axes; vascular and nutrient delivery systems (scaffolding & plumbing, c.f. *ex cranio* brains)

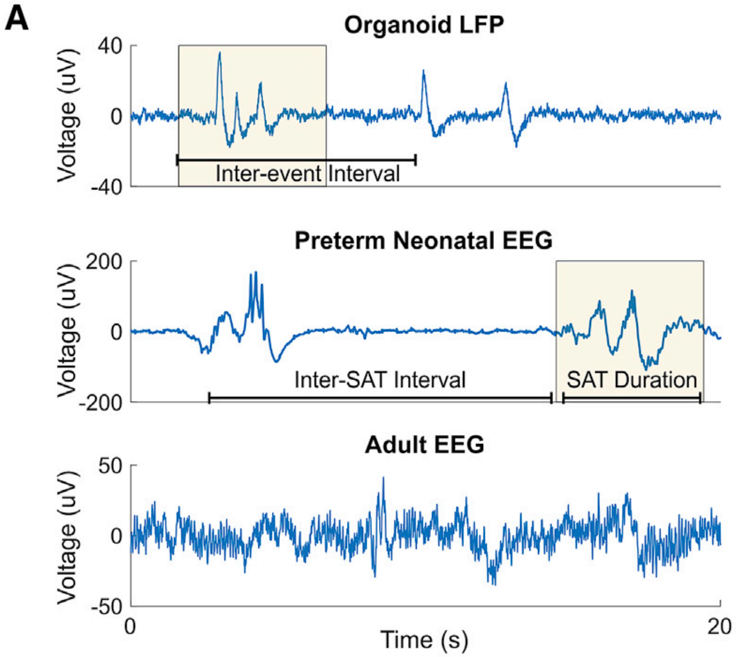
Trujillo et al (2019) *Cell Stem Cell*

Hostiuc et al (2019) *Regen Therapy*

neural signatures of consciousness

- number of neurons?
- any particular neuronal type (relevant for chimeras)?
- ’ignition’ of a ‘global workspace’
- widespread recurrent information flow
- ‘complexity’ or ‘integrated information’
- high-frequency low-amplitude EEG
- 1/f activity profile (possibly)
- …. among others ….





Trujillo et al (2019) *Cell Stem Cell*

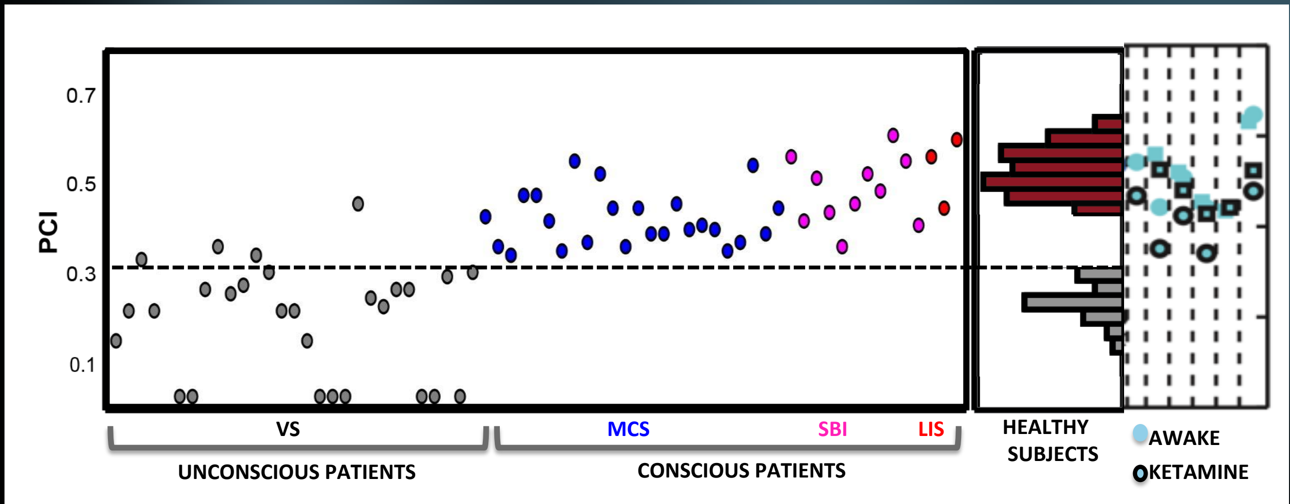
consciousness in organoids?

- possibility of false positives
- presence of brain-like features may encourage attribution of consciousness

- possibility of false negatives
- differences from mature brains, and isolation from environment, may prevent attribution of consciousness

an ‘architecture agnostic’ signature

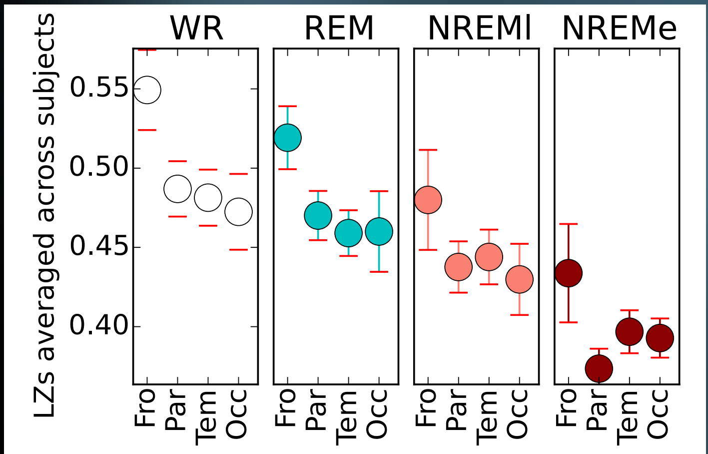
- algorithmic complexity of brain response (EEG) to a TMS pulse: the perturbational complexity index (PCI)
- loosely operationalises the ‘integrated information theory’ of consciousness (IIT) (but note, this is is *not* a consenus theory)



Tononi et al (2016) *Nat Rev Neurosci*
Casali et al (2013) *Sci Trans Med*

PCI for organoids?

- utilise e.g. optogenetic stimulation and calcium imaging (already some adaptations of PCI for cortical slices)
- use non-perturbational approximations (e.g., brain entropy as measured by Lempel-Ziv complexity)
- accentuates the problems of baseline and common scale



Schartner et al (2017) *Neurosci Consc*

Bayne, Seth, Massimini (2020) *Trends Neurosci*

Lavazza & Massimini (2018) *J Med Ethics*

challenges for assessing organoid consciousness

- lack of consensus theory on the sufficient mechanisms for consciousness
- lack of baselines/scales to allow meaningful interpetation of architecture-agnostic signatures (such as PCI)
- lack of any behavioural verification / validation to underwrite generalisation of existing measures

Shepherd (2018) *J Med Ethics*
Bayne, Seth, Massimini (2020) *Trends Neurosci*

addressing the challenges

- psychological architecture: instead of looking for bare awareness (based on one theory), look for neural signatures of psychological states of ethical significance (e.g., affective/ evaluative) that share features across theories.

- incremental generalisation: progressively broaden PCI out from healthy human subjects to DoC patients, non-human animals, and ultimately organoids

philosophical implications

- the possibility of islands of awareness, such as conscious organoids, puts pressure on (exernalist) theories of consciousness that propose the (synchronic or diachronic) necessity of sensorimotor interactions

Bayne, Seth, Massimini (2020) *Trends Neurosci*

preventative ethics

- organoids combine large-scale production with epistemic uncertainty (cannot use behaviour to infer presence of pain or suffering)
- while organoid consciousness may be unlikely, it is probably less unlikely than machine consciousness
- possibility of combining organoid and chimera research
- this suggests a need for preventative ethics

ARTICLE

https://doi.org/10.1038/s41586-019-1099-1

Restoration of brain circulation and cellular functions hours post-mortem

Zvonimir Vrselja[1,2,18], Stefano G. Daniele[1,2,3,18], John Silbereis[1,2], Francesca Talpo[1,2,4], Yury M. Morozov[1,2], André M. M. Sousa[1,2], Brian S. Tanaka[5,6,7], Mario Skarica[1,2], Mihovil Pletikos[1,2,8], Navjot Kaur[1,2], Zhen W. Zhuang[9], Zhao Liu[9,10], Rafeed Alkawadri[6,11], Albert J. Sinusas[9,10], Stephen R. Latham[12], Stephen G. Waxman[5,6,7] & Nenad Sestan[1,2,13,14,15,16,17]*

The brains of humans and other mammals are highly vulnerable to interruptions in blood flow and decreases in oxygen levels. Here we describe the restoration and maintenance of microcirculation and molecular and cellular functions of the intact pig brain under ex vivo normothermic conditions up to four hours post-mortem. We have developed an extracorporeal pulsatile-perfusion system and a haemoglobin-based, acellular, non-coagulative, echogenic, and cytoprotective perfusate that promotes recovery from anoxia, reduces reperfusion injury, prevents oedema, and metabolically supports the energy requirements of the brain. With this system, we observed preservation of cytoarchitecture; attenuation of cell death; and restoration of vascular dilatory and glial inflammatory responses, spontaneous synaptic activity, and active cerebral metabolism in the absence of global electrocorticographic activity. These findings demonstrate that under appropriate conditions the isolated, intact large mammalian brain possesses an underappreciated capacity for restoration of microcirculation and molecular and cellular activity after a prolonged post-mortem interval.

Many mammalian species have large, energy-demanding brains that are highly susceptible to anoxia and cessation of blood flow[1–3]. Studies in both humans and experimental animals have shown that oxygen stores, global electrical activity, and consciousness are lost within seconds of interrupted blood flow, while glucose and ATP stores are depleted within minutes[4–8]. Unless perfusion is quickly restored, multiple deleterious mechanisms lead to widespread membrane depolarization, loss of ionic homeostasis, mitochondrial dysfunction, and excitotoxic accumulation of glutamate[9,10]. The convergence of these factors has been widely proposed to initiate a progressive, and largely irreversible, cascade of apoptosis, necrosis, and axonal damage[4–9].

However, several observations have questioned the inevitability of neural cell death minutes, or even hours, after cessation of brain perfusion. First, tissue specimens with sufficient viability for cell and organotypic slice cultures[11,12,13], as well as for electrophysiological recordings[14], have been taken from human and other mammalian brains hours after death. Second, mitochondria remain functional for up to 10 h post-mortem in human cerebral cortical tissue[15]. Third, in cats and macaques, 1 h of complete global ischaemia can be followed by neuronal, electrophysiological, and metabolic recovery after reperfusion[16–19]. Last, full neurological recovery from prolonged asystole has been reported in humans with hypothermia[20], and recent clinical findings have suggested that thrombectomies performed up to 16 h after an ischaemic insult can result in favourable patient outcomes[21]. These data suggest that the initiation and duration of cell death after anoxia or ischaemia may span a longer temporal interval than is currently appreciated, allowing for a multifaceted intervention that could halt the progression of damaging cellular programs initiated by the global insult. Therefore, we postulate that, under appropriate conditions, certain molecular and cellular functions in the large mammalian brain may retain at least partial capacity for restoration after a prolonged post-mortem interval (PMI).

To test this hypothesis, we developed a surgical procedure, perfusate, and custom pulsatile-perfusion device that can restore and maintain microcirculation and cellular viability in the large mammalian brain under ex vivo normothermic conditions (37 °C) after an extended PMI. This system is herein referred to as Brain*Ex* (B*Ex*). To determine whether restoration and maintenance of cell viability is possible, we engineered a haemoglobin-based, acellular, echogenic, and non-coagulative cytoprotective B*Ex* perfusate. In order to develop all aspects of this technology, we reasoned that a prudent approach would be to utilize post-mortem brain specimens from USDA-regulated food processing facilities, which would otherwise be discarded. Therefore, we applied this technology to the isolated, and largely ex cranio, brains of 6–8-month-old pigs (*Sus scrofa domesticus*) 4 h post-mortem. Using this approach, we observed attenuation of cell death and preservation of anatomical and neural cell integrity. We also found that specific cellular functions were restored, as indicated by vascular and glial responsiveness to pharmacological and immunogenic interventions, spontaneous synaptic activity, and active cerebral metabolism in the absence of global brain activity.

These findings show that, with appropriate interventions, the large mammalian brain retains an underappreciated capacity for normothermic restoration of microcirculation and certain molecular and cellular functions multiple hours after circulatory arrest. In addition, this platform could offer investigators the opportunity to conduct prospective,

[1]Department of Neuroscience, Yale School of Medicine, New Haven, CT, USA. [2]Kavli Institute for Neuroscience, Yale School of Medicine, New Haven, CT, USA. [3]Medical Scientist Training Program (MD-PhD), Yale School of Medicine, New Haven, CT, USA. [4]Department of Biology and Biotechnology L. Spallanzani, University of Pavia, Pavia, Italy. [5]Center for Neuroscience and Regeneration Research, Yale School of Medicine, New Haven, CT, USA. [6]Department of Neurology, Yale School of Medicine, New Haven, CT, USA. [7]Rehabilitation Research Center, VA Connecticut Healthcare System, West Haven, CT, USA. [8]Department of Anatomy and Neurobiology, Boston University School of Medicine, Boston, MA, USA. [9]Section of Cardiovascular Medicine, Department of Internal Medicine, Yale School of Medicine, New Haven, CT, USA. [10]Department of Radiology and Biomedical Imaging, Yale School of Medicine, New Haven, CT, USA. [11]Department of Neurology, University of Pittsburgh, Pittsburgh, PA, USA. [12]Interdisciplinary Center for Bioethics, Yale University, New Haven, CT, USA. [13]Department of Genetics, Yale School of Medicine, New Haven, CT, USA. [14]Department of Psychiatry, Yale School of Medicine, New Haven, CT, USA. [15]Department of Comparative Medicine, New Haven, CT, USA. [16]Yale Child Study Center, Yale School of Medicine, New Haven, CT, USA. [17]Program in Cellular Neuroscience, Neurodegeneration and Repair, Yale School of Medicine, New Haven, CT, USA. [18]These authors contributed equally: Zvonimir Vrselja, Stefano G. Daniele. *e-mail: nenad.sestan@yale.edu

Vrselja *et al* (2019) *Nature*

COMMENT

RODOLFO BUHRER/REUTERS

Pigs raised for food production are exempt from welfare laws governing how research animals are treated.

Part-revived pig brains raise ethical quandaries

Researchers need guidance on animal use and the many issues opened up by a leap in brain restoration, urge **Nita A. Farahany**, **Henry T. Greely** and **Charles M. Giattino**.

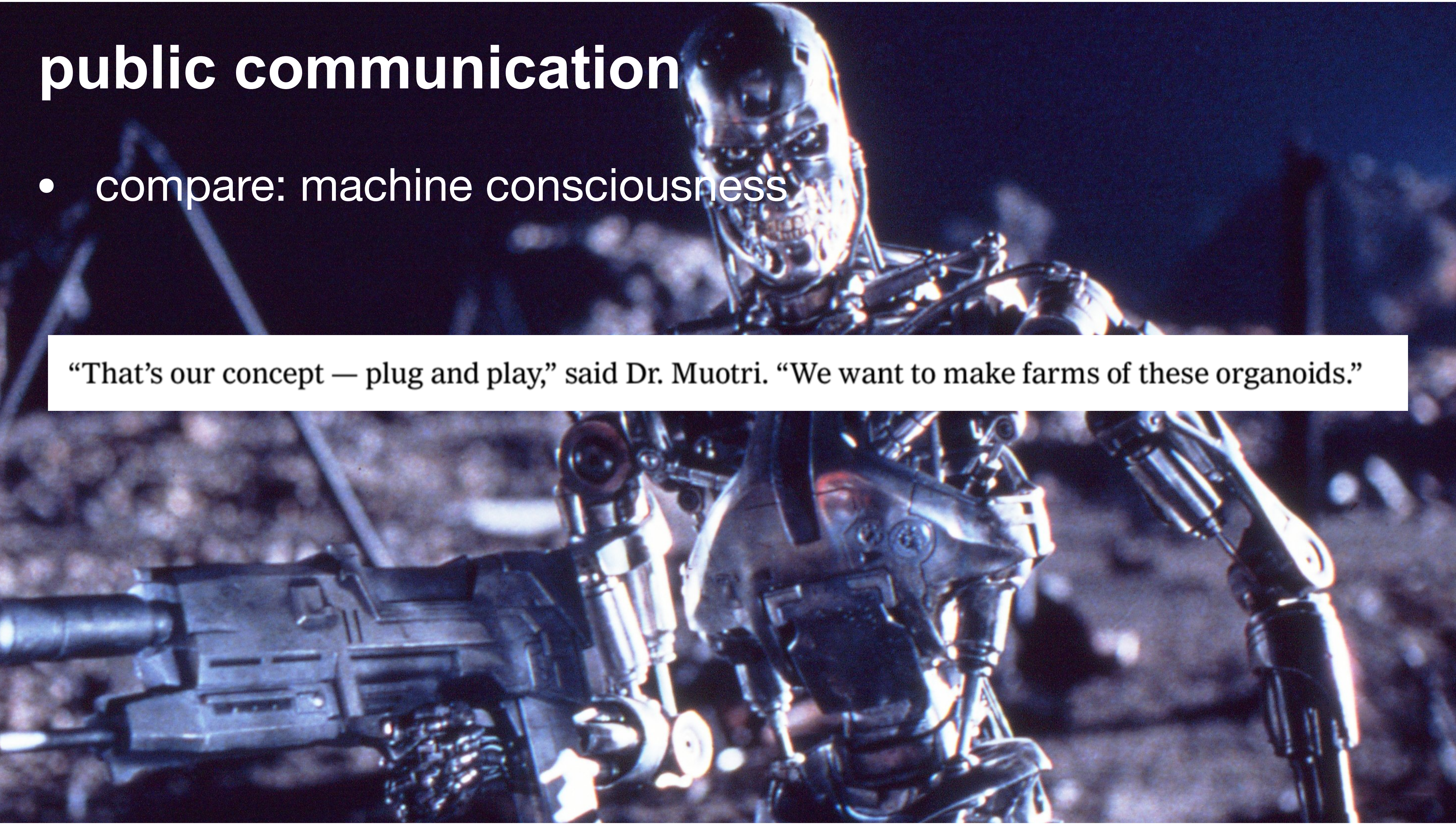
Scientists have restored and preserved some cellular activities and structures in the brains of pigs that had been decapitated for food production four hours before. The researchers saw circulation in major arteries and small blood vessels, metabolism and responsiveness to drugs at the cellular level and even spontaneous synaptic activity in neurons, among other things. The team formulated a unique solution and circulated it through the isolated brains using a network of pumps and filters called Brain*Ex* (see page 302). The solution was cell-free, did not coagulate and contained a haemoglobin-based oxygen carrier and a wide range of pharmacological agents.

The remarkable study, published in this week's *Nature*[1], offers the promise of an animal or even human whole-brain model in which many cellular functions are intact. At present, cells from animal and human brains can be sustained in culture for weeks, but only so much can be gleaned from isolated cells. Tissue slices can provide snapshots

Farhany *et al* (2019) *Nature*

public communication

- compare: machine consciousness



“That’s our concept — plug and play,” said Dr. Muotri. “We want to make farms of these organoids.”

summary

- inadvertently creating organoid (or chimera) consciousness would be ethically catastrophic
- the possibility, however outlandish, must be taken seriously
- use architecture-agnostic adaptations of human/mammal-validated measures as a starting point
- consider a range of signatures, rather than a single indicator
- as with disorders of consciousness, this challenge underlines the practical relevance of consciousness research (& importance of funding it!)

Michel *et al* (2019) *Nature Human Behaviour*

key references

Bayne, T., Seth, A.K., and Massimini, M. (2020). Are there islands of awareness? *Trends in Neurosciences*.

Bayne, T., Seth, A.K., and Massimini, M. (2020). From complexity to consciousness: Reply to Cecconi et al. *Trends in Neurosciences*.

Schartner, M., Pigorini, A., Gibbs, S., Sarasso, S., Barnett, L.C., Nobili, L., Massimini, M., Seth, A.K., and Barrett, A.B. (2017). Global and local complexity of intracranial EEG decreases during NREM sleep. *Neuroscience of Consciousness* 3(1):22

Casali, A. et al. (2013). A theoretically based index of consciousness independent of sensory processing and behaviour. *Sci Trans Med* 5(198):198ra105

Lavazza, A., & Massimini, M. (2018). Cerebral organoids: ethical issues and consciousness assessment. *J Med Ethics* 44:6-10

islands of awareness

Anil K Seth

Sackler Centre for Consciousness Science
University of Sussex
CIFAR Program on Brain, Mind, and Consciousness

NASEM Committee: Organoids and Chimeras

University of Sussex
Sackler Centre for Consciousness Science

Please cite this article in press as: Bayne et al., Are There Islands of Awareness?, Trends in Neurosciences (2019), https://doi.org/10.1016/j.tins.2019.11.003

Trends in Neurosciences

CellPress REVIEWS

Opinion

Are There Islands of Awareness?

Tim Bayne,[1,6,7,*] Anil K. Seth,[2,3,6,7] and Marcello Massimini[4,5,6,7]

Ordinary human experience is embedded in a web of causal relations that link the brain to the body and the wider environment. However, there might be conditions in which brain activity supports consciousness even when that activity is fully causally isolated from the body and its environment. Such cases would involve what we call islands of awareness: conscious states that are neither shaped by sensory input nor able to be expressed by motor output. This opinion article considers conditions in which such islands might occur, including *ex cranio* brains, hemispherotomy, and in cerebral organoids. We examine possible methods for detecting islands of awareness, and consider their implications for ethics and for the nature of consciousness.

Highlights

Awareness may persist in fully disconnected cortical islands.

We identify both natural and artificial examples of potential islands of awareness.

Detecting islands of awareness poses difficult but often addressable challenges.

The possibility of islands of awareness raises important ethical and legal issues.

The discovery of islands of awareness would have important implications for debates about the nature of consciousness.

The Challenge of Islands of Awareness

Consciousness is first and foremost a property of living organisms – organisms that are embodied and embedded in environments. The contents of consciousness are shaped by the sensory stimuli received by the brain, and those contents in turn give rise to behaviours that prompt us to attribute consciousness to an organism. However, there are conscious states in which the transfer of information between the world and the brain is massively reduced, with the result that the brain (or parts thereof) becomes disconnected from its environment. In some conditions, disconnection is partial, so that some form of either input and/or output is retained. In other conditions, the disconnection is complete, so that the brain (or parts thereof) becomes fully isolated from its environment.

What happens to consciousness when the brain becomes disconnected from its environment? Can it support **islands of awareness** (see Glossary), or does consciousness require the presence of (high bandwidth?) interaction between the brain and its environment? This question has long fascinated philosophers, but recent developments in neuroscience, neurosurgery, and neuroengineering now extend the scope of this discussion beyond the philosopher's armchair and out into the laboratory and clinic.

We address three issues raised by the possibility of islands of awareness. The first concerns their nature and distribution. Under what conditions might such islands arise? What forms might they take? How common might they be? A second issue concerns the detection of islands of awareness. Might current methods for detecting consciousness be applicable to islands of awareness, or will we need new tools for identifying consciousness in disconnected brains? A third issue concerns the implications of islands of awareness. What ethical implications might such islands have, and what might they tell us about the nature of consciousness?

We address these issues by considering three conditions in which islands of awareness might be thought to occur: *ex cranio* brains; the neurosurgical procedure of hemispherotomy; and **cerebral organoids**. Although these three cases are by no means the only cases that could be considered here – for example, one might also consider whether islands of awareness could occur *in utero* [1] – we focus on them here because they highlight the issues raised by islands of awareness with particular force and urgency.

From Partial Disconnection to Complete Disconnection

Before we turn to genuine islands of awareness, we begin with cases of merely partial disconnection. Clinical neurology offers a rich repertoire of cases to consider here, for structural lesions can cause the brain to become disconnected on either the input or the output side without loss of consciousness.

Starting from the input side, we know that consciousness can be preserved in the absence of afferent activity from peripheral receptors and nerves. For example, acquired blindness is a condition in which

[1]School of Philosophical, Historical, and International Studies, Monash University, Melbourne, Australia

[2]School of Engineering and Informatics, University of Sussex, Brighton, UK

[3]Sackler Centre for Consciousness Science, University of Sussex, Brighton, UK

[4]Department of Biomedical Clinical Sciences, University of Milan, Milan, Italy

[5]IRCCS, Fondazione Don Gnocchi, Milan, Italy

[6]Canadian Institute for Advanced Research, Azrieli Program in Brain, Mind and Consciousness, Toronto, Canada

[7]These authors contributed equally to this paper

*Correspondence: timothy.bayne@monash.edu

Please cite this article in press as: Bayne et al., Are There Islands of Awareness?, Trends in Neurosciences (2019), https://doi.org/10.1016/j.tins.2019.11.003

patients lose sight but retain the capacity for imagery, visual dreaming, and vivid hallucination [2]. A corresponding dissociation can also occur in audition [3]. We also know that direct cortical stimulation within the appropriate parameters can elicit visual, auditory, tactile, and interoceptive experiences while bypassing subcortical sensory pathways [4] (see also [5] for a recent study in which perceptual discrimination was induced in rodents through optogenetic stimulation of visual cortex). Although we are not aware of any case in which all sensory pathways have been structurally severed without concurrent motor nerve impairment, approximations to a severe multimodal disconnection can occur in the late stages of multiple sclerosis or mitochondrial disorders [6].

On the motor side, a classic example of partial disconnection is locked in syndrome (LIS), which can occur as the result of a ventral pontine lesion severing of all motor fibres except for the third cranial nerve (which drives vertical eye movements and blinking). Despite almost complete motor disconnection, LIS patients can be fully communicative and are undoubtedly conscious. However, in some patients the neurons of the third cranial nerve are also impaired and the patient is unable to produce any detectable motor output [7,8]. In such patients the brain is completely disconnected on the output side, but there is every reason to think that consciousness has been retained despite the loss of even the last channel of motor output.

There are also cases in which the brain becomes disconnected from its environment in both input and output terms, although the disconnection is not always absolute and is often reversible. One form of disconnection occurs in dreaming, when changes in neuromodulation result in cortical gating of sensory inputs and in inhibition of motor neurons [9]. More comprehensive disconnection can occur under the influence of the dissociative anaesthetic ketamine, for at certain doses ketamine blocks both exteroceptive and interoceptive input, allowing patients to undergo invasive surgical procedures. But despite disconnection and profound unresponsiveness ketamine does not always extinguish consciousness, and can instead induce vivid and sometimes terrifying experiences [10].

Although dreaming and ketamine involve consciousness in the context of sensory and motor disconnection, in both cases the disconnection is functional rather than structural. However, reversible disconnection can also result from structural factors. The rare case of a conscious patient in whom concurrent conditions resulted in complete sensory (visual, auditory, and tactile modalities) and motor disconnection was recently reported [11]. Sensory–motor disconnection with preserved awareness can also be observed in extreme cases within the spectrum of acute inflammatory polyneuropathy, such as fulminant Guillain–Barré syndrome (GBS) [12]. In some GBS patients, a complete paralysis extending to cranial nerves can be accompanied by a severe (albeit not complete) blockage of multiple sensory nerves [13–15]. In the initial stages of the condition patients are clearly conscious and can communicate through residual movements. Many report vivid dreams and intense hallucinations, some of which resemble the hallucinations that are induced by weightlessness in astronauts or experienced by those in sensory deprivation tanks [16]. However, all movements are quickly lost and communication becomes impossible. Unresponsiveness can be so deep, even extending to absence of cranial reflexes, that the condition can mimic brain death. However, magnetic resonance imaging (MRI) shows no alteration in the central nervous system and the limited electroencephalography (EEG) available in this condition shows either normal wakeful patterns or mild slowing [13]. In some cases, patient awareness can also be demonstrated by preserved EEG and metabolic responses to auditory stimulation [17] and verbal commands [18]. As the condition persists, the sleep–wake cycle breaks down and the EEG becomes more difficult to interpret. Patients who gradually recover motor control and functional communication tend to be confused and amnestic but some recall having been vividly conscious, albeit in an altered state, while completely paralysed in their acute illness [19,20].

Islands of Awareness

Although dreaming, ketamine and GBS can each involve (relatively) complete disconnection from the environment, in each case the brain retains some capacity for being reintegrated into its environment. Indeed, it is this capacity that enables the retrospective reports which – together with neuroimaging data – supports the inference that consciousness has been retained. However, it is interesting to ask

Glossary

Hemispherectomy/otomy: procedure used for the treatment of certain severe cases of epilepsy in which an entire brain hemisphere is either surgically removed from the cranium and discarded (hemispherectomy), or the connections between the hemisphere and the rest of the brain are cut but the hemisphere itself is left *in situ* (hemispherotomy).

Islands of awareness: conscious stream (or system) whose contents are not shaped by sensory input from either the external world or the body and which cannot be expressed via motor output.

Cerebral organoids: stem-cell-derived laboratory-grown structures that self-organise into three dimensions with cellular and network features resembling certain aspects of the developing human brain.

Perturbational complexity index (PCI): technique in which TMS is used to perturb the cortex, and EEG is then used to measure the electrocortical responses to that perturbation. The algorithmic complexity (information) inherent in that electrocortical response is taken to be an indicator of consciousness.

whether consciousness might also occur even in systems that lack any sensory or motor connections to the body or environment. We call such centres of consciousness 'islands of awareness'. An island of awareness can be thought of as a limiting case of the kinds of sensory and motor disconnections that we have considered above (Figure 1). In each of those cases disconnection is merely partial, for at least some sensorimotor pathways between the brain and its environment are retained. A genuine island of awareness, however, has no sensorimotor interaction with the body that supports it, nor with the environment that surrounds it. It is also important to distinguish islands of awareness from instances of covert consciousness, in which consciousness is not manifested in outward behaviour, and for which there is now significant evidence in at least some behaviourally unresponsive patients who have emerged from coma [21,22]. Islands of awareness are more profoundly isolated: their experiences are causally isolated from both the body and the environment, and in both motor and sensory respects.

In what kinds of situations might islands of awareness occur? It is certainly possible that they might occur in the context of severe multifocal brain injury (Box 1). However, we will focus here on conditions that have not previously been considered in connection with islands of awareness – conditions that have been made possible by advances in neurosurgery and neuroengineering.

We begin with a study by Vrselja and colleagues [23], in which intact pig brains were extracted up to 4 h postmortem and then connected to a system called BrainEx, which delivered nutrients and oxygen to brain cells. Vrselja and colleagues were able to restore and maintain microcirculation as well as molecular and cellular functions of neurons under *ex vivo* conditions for several hours, and without any sensory input or motor output. Significantly, they observed spontaneous synaptic activity in these *ex cranio* brains. Although there was no evidence of global brain activity and no EEG response, that fact might be explained by their use of a preservative solution that inhibited neural activity. Previous research in which guinea pig brains were isolated and perfused *in vitro* in hypothermic conditions but without pharmacological blocking agents has demonstrated preservation of electrical responses across multisynaptic circuits [24] as well as synaptic plasticity [25].

While these data do not provide direct evidence for awareness in a 'naked brain', [24], they certainly raise that possibility. Suppose that the experiment conducted by Vrselja and colleagues were to be repeated without an explicit neural inhibitor and under normothermic conditions. If organised patterns of spontaneous neural activity were to be observed in this situation, the question of whether an island of awareness was present would immediately arise (just as it immediately arises in the case of fully disconnected brains in some GBS patients).

Might islands of awareness occur even in an isolated cortical hemisphere? The relatively rare neurosurgical procedure of **hemispherotomy**, our second case, involves the disconnection of a damaged hemisphere in order to treat children with severe cases of refractory epilepsy [26]. Hemispherotomy aims at maximal disconnection of white matter pathways linking the pathological hemisphere to the brainstem, thalamus, and contralateral hemisphere. The damaged hemisphere is, however, left inside the cranial cavity with vascular connectivity intact. In the related procedure of hemispherectomy, the damaged hemisphere is first disconnected and then removed from the brain entirely. Typical peri-insular hemispherotomy involves severing the corona radiata, resection of the temporal lobe, complete section of the corpus callosum, subfrontal and temporal stem disconnection, and undercutting or resection of the insular cortex [27]. While neural disconnection is usually assumed to be complete, some residual commissural and central connections via the hypothalamus and optic chiasm may remain (Michael Carter, personal communication). It is clear that the intact and properly connected hemisphere supports awareness (for patients are clearly conscious), but could the disconnected hemisphere also support awareness? If so, then there would be an island of awareness in the disconnected hemisphere.

Little is known about the consequence of this radical deafferentation on neural activity in the disconnected hemisphere. One recent study performed intraoperative electrocorticography and found reduced broadband spectral power in regions of disconnected cortex surrounding the pathological

Please cite this article in press as: Bayne et al., Are There Islands of Awareness?, Trends in Neurosciences (2019), https://doi.org/10.1016/j.tins.2019.11.003

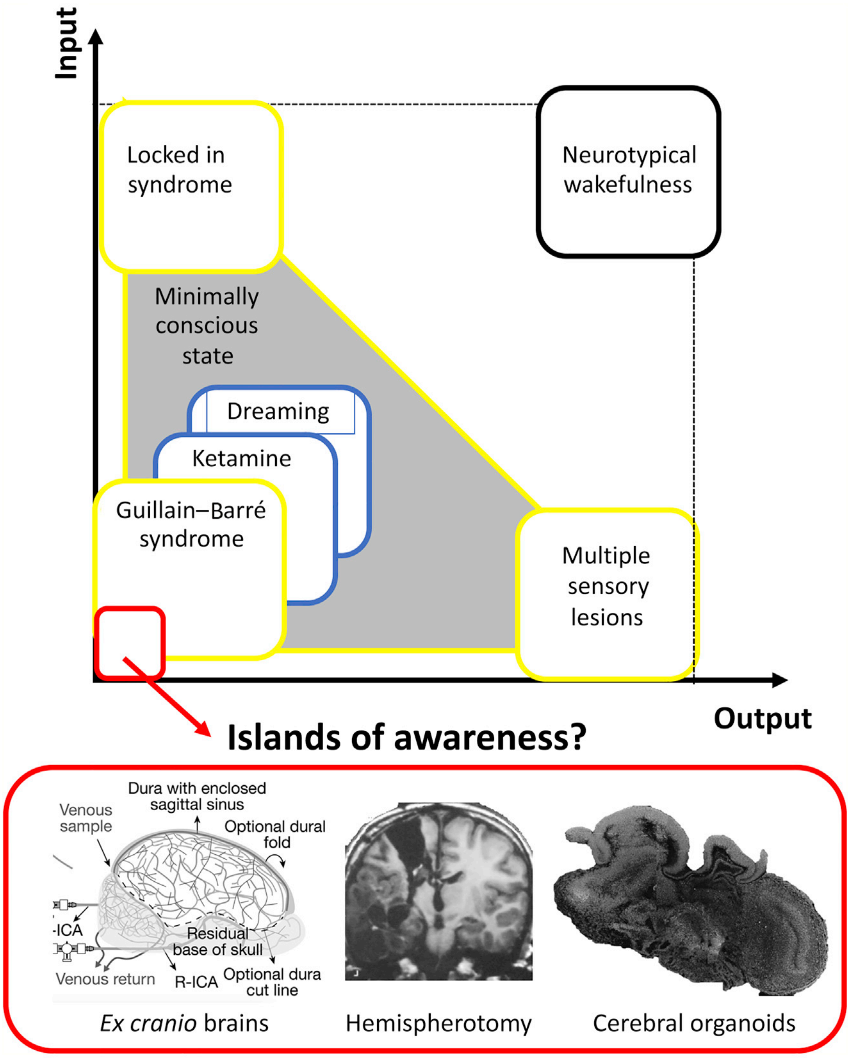


Figure 1. Schematic Representation of Some Possible Instances of Islands of Awareness in Relation to Other Conditions of Consciousness.

The graph represents different conditions as a function of their approximate level of sensory and motor connectedness. The black boundary defines the high level of connectedness typical of healthy awake humans. The blue boundaries include conscious states characterised by forms of disconnection that are functional and reversible (e.g., dreaming, and hallucinations during ketamine anaesthesia). The yellow boundaries include pathological conditions in which consciousness might be preserved in spite of various degrees of structural disconnection. They encompass cases of pure motor disconnection (e.g., locked in syndrome); cases of pure sensory disconnection (e.g., multiple concurrent lesions of sensory systems); cases in which multifocal brain injury may affect both motor and sensory systems to a variable extent (e.g., the large grey area of the minimally conscious state, MCS); and situations that approximate complete, albeit reversible, sensory and motor disconnection (e.g., Guillain –Barré syndrome). The red boundaries identify conditions of complete, irreversible structural disconnection. As depicted in the inset below, they include *ex cranio* brains, disconnected hemispheres post-hemispherotomy and cerebral organoids. *Ex cranio* brains [23], hemispherotomy [26], cerebral organoids [31]; reproduced with permission. Abbreviation: R-ICA, right internal carotid artery.

Box 1. Islands of Awareness in Disorders of Consciousness

Following their emergence from the comatose state, many patients who have suffered from serious brain injury spend a period of time in the vegetative state (also known as the unresponsiveness wakefulness syndrome), in which they fail to produce any behavioural signs of consciousness. However, neuroimaging and EEG active paradigms suggest that roughly 15% of these patients can modulate their brain activity to verbal command, which is indicative of (covert) consciousness [54].

It is possible, however, that these figures misrepresent the true prevalence of covert consciousness in behaviourally nonresponsive patients, for the multifocal nature of traumatic brain injury often results not only in motor and sensory impairments, but also in cognitive and attentional impairments. A significant proportion of these patients might fail both overt and covert tests of consciousness not because they are unconscious, but simply because they are unable to process or cognitively engage with sensory stimuli [11,55]. Such cases highlight the fact that intensive care medicine is likely already creating instances of disconnected consciousness, with the precise nature of the disconnection in any one case due to both structural and functional factors. It is also possible that the behavioural fragments that are sometimes seen in these conditions – such as the utterance of a single word [56–58] – are manifestations of transient conscious experiences, rather than merely being the products of unconscious motor routines. Addressing the challenges posed by genuine islands of awareness, in which the presence of consciousness must be inferred without reliance on any intact sensory or motor pathway, might therefore facilitate the development of increasingly sensitive methods for detecting covert consciousness in post-comatose patients.

tissue [28]. This study also found increased local functional connectivity in this apparently preserved cortex, using mutual information analysis approaches. Intriguingly, an earlier functional MRI (fMRI) study reported surprising and unexplained task-related blood oxygenation level dependent (BOLD) activations in a disconnected left hemisphere, following left hemispherotomy in a child with Rasmussen's syndrome [29].

Additional evidence about electrophysiological activity in isolated cortex has been provided by recordings in cortical volumes of various sizes (cortical slabs), which are deafferented from the rest of the brain through a white matter undercut. Timofeev and colleagues [30] showed that small isolated cortical volumes (10×6 mm) can sustain sporadic depolarizing events at a frequency of 0.03–0.1 Hz. Notably, the pattern evolves toward a sleep-like slow oscillation in the delta range ($\sim$1 Hz) if the volume of the cortical slab is larger (30×20 mm, roughly corresponding to a cortical gyrus), allowing for more recurrent excitatory activity.

This stereotypical pattern resembles the slow oscillations normally observed in intact brains during dreamless sleep and may result from a lack of input from the thalamus and subcortical activating systems to the isolated cortex. This raises the possibility that replacing that input – for example, by direct electrochemical stimulation of cortical neurons or long-term homeostatic processes – might restore adequate levels of cortical excitability, and with this perhaps also some form of consciousness (Box 2).

The third case in which islands of awareness might occur involves cerebral organoids. These are laboratory-made 3D structures derived from stem cells that display various features of the developing human brain [31]. Cerebral organoids are sometimes called 'mini brains', although it is arguable that this overstates their similarity to normal brains. The primary use of cerebral organoids has been as laboratory models of neurodevelopmental disorders such as Zika-virus-induced microcephaly, but prospective applications could encompass a wide range of neurological conditions [32]. Progress in cerebral organoid development has been rapid, with recent organoids demonstrating mature neurons and established network structures. Assessment of neural activity using calcium imaging and high-density silicon microelectrodes has revealed spiking activity [33] and complex oscillatory waves resembling some features of human pre-term EEG [34]. While cerebral organoids still lack a well-defined neuroanatomical axis, as well as vascular and nutrient-delivery systems, it is not unreasonable to suppose that near- or mid-term developments will deliver organoids displaying substantial structural and functional similarities to developing human brains.

Please cite this article in press as: Bayne et al., Are There Islands of Awareness?, Trends in Neurosciences (2019), https://doi.org/10.1016/j.tins.2019.11.003

Box 2. What Is an Enabling Factor?

An important issue raised by many disconnection scenarios concerns the role that subcortical systems places in the generation of consciousness. These systems are often described as enabling factors for consciousness, but that phrase can be understood in two ways: causally or constitutively. In the causal sense, an enabling factor makes a causal contribution to the target process in the way in which a lightning strike might cause a forest fire, such that its causal contribution could in principle be provided by another process (say, a burning cigarette). If subcortical systems are enabling factors in this sense then it is at least possible that their causal role could be replaced by some other factor (say, electrochemical stimulation), and cortical processes alone might suffice for consciousness. In the constitutive sense, an enabling factor is a crucial component of the minimal neural substrate of consciousness. Its role is not to cause cortical systems to enter a state in which they generate consciousness – rather, subcortical activation is itself a component of the neural basis of consciousness. On one version of this account, subcortical activity does not explain why consciousness has the particular contents that it does (it is not part of the differentiating neural correlates of consciousness), but it does play an essential role in explaining why consciousness of any kind occurs (it is a nondifferentiating correlate of consciousness). If subcortical systems turn out to be enabling factors in the constitutive sense, then islands of awareness would not be possible in a disconnected cortical hemisphere, but would still occur in the context of an *ex cranio* brain in which cortical activity was appropriately integrated with subcortical activity. Alternatively, different subcortical structures might turn out to be enabling factors of different kinds. For example, brainstem activating systems and midline thalamic nuclei, modulating the excitability of cortical neurons, might have a casual role, whereas high-order thalamic nuclei and (say) the claustrum, granting tight structural integration among distant cortical areas, might be constitutive.

Assessing the possibility of consciousness in cerebral organoids faces challenges that do not apply to the other cases we have discussed. On the one hand, organoids develop as intact wholes, without having ever had any causal interaction with the external world. This might be thought to mitigate against consciousness if such connectivity is constitutively necessary for developing consciousness (Box 3). At the same time, the fact that organoids develop 'naturally' and do not suffer from radical disruption to their neural structure (in the way that *ex cranio* brains and hemispherotomy patients do) might be thought to argue in favour of the possibility of conscious organoids. In short, the question of consciousness in cerebral organoids remains open. Figure 1 summarises the various cases that we have examined thus far, including both cases of disconnected consciousness (in which connection can sometimes be regained) and what we term islands of awareness, in which neither sensory input nor motor output can be achieved.

Detecting Islands of Awareness

Our capacity to tell whether another creature is conscious ordinarily relies on inferences from behaviour, and by definition islands of awareness have no motor output. Although methods have been developed for detecting consciousness in behaviourally nonresponsive patients [21,35,36], most of these methods require intact sensory pathways and thus they cannot be applied to the cases that we are considering here. How then might genuine islands of awareness be detected?

One interesting possibility is that consciousness can be detected by assessing causal interactions within the brain, even when reciprocal interactions with its surrounding environment are completely interrupted. A practical way of probing the internal causal structure of the brain involves a perturb-and-measure approach through a combination of cortical stimulation and neuroimaging [37]. An example of this method is provided by the **perturbational complexity index (PCI)** in which the cerebral cortex is first stimulated by transcranial magnetic stimulation (TMS) and then EEG is used to measure the complexity of the cause–effect chain of neural activations triggered within the brain [38]. PCI has proven effective in detecting disconnected awareness during dreaming and ketamine anaesthesia [39] and proven able to identify conscious patients who are minimally responsive, or fully unresponsive, following severe brain injury [40]. In spite of its high accuracy, the general applicability

Please cite this article in press as: Bayne et al., Are There Islands of Awareness?, Trends in Neurosciences (2019), https://doi.org/10.1016/j.tins.2019.11.003

Box 3. Validating Measures of Consciousness

Validating novel measures of consciousness that can be used to detect consciousness in difficult cases – such as those involving putative islands of awareness – raises deep methodological challenges [59–61]. The standard approach to validating a novel measure is to correlate the presence/absence of that measure with a given pretheoretical measure of consciousness, such as behavioural responsiveness. However, this approach faces problems when making inferences in cases where the pretheoretical measure by definition cannot be obtained, such as in disconnected patients. To address this problem, a measure's validation can be extended to fit also cases of disconnection; for example, by relying on delayed report upon awakening from dreaming and ketamine dissociation. This validation strategy has already been utilised to allow for improved inference in brain-injured patients with sensory and motor disconnection [39]. More generally, confidence in novel measures can be gained on the basis of their overall convergence with multiple pretheoretical and previously validated measures of consciousness [62–64]. Confidence is further increased when such measures can be demonstrated to operationalise a particular theory of consciousness, to the extent that that theory shows general explanatory and predictive power.

A second challenge is that measures of consciousness validated or otherwise applied in adult human beings might not transfer to (say) infants or nonhuman animals. This worry applies with particular force to cerebral organoids, because the kinds of neural activity that support consciousness in an organoid might differ in important respects from those that support consciousness in human brains. Relying on similarities between organoid activity and human brain activity might lead both to the overattribution of conscious states to organoids (false positives; [33]) and, simultaneously, to the underattribution of conscious states (false negatives) given that organoid consciousness might involve very different kinds of activity patterns. Having said this, if theoretically grounded and empirically robust measures applicable in humans are applied with similar results in nonhuman cases, this should be taken as highly suggestive of consciousness in such cases.

The history of science has repeatedly encountered the problem of validating measures of incompletely understood phenomena, when accurate measurement is in turn essential for reaching a satisfying scientific understanding of the target phenomenon. For example, the development of reliable thermometers faced and overcame these problems in catalysing a physical explanation of heat [65]. Similarly, careful and incremental extension of novel measures of consciousness to difficult cases could similarly catalyse a deeper physical understanding of the nature of consciousness that could then in turn further validate these novel measures.

of PCI presents practical challenges, including (for hemispherotomy) the risks posed by electrical stimulation of an epilepsy-prone brain.

There are, however, ways to estimate the complexity of neural dynamics without applying cortical perturbations. For example, measures of the algorithmic complexity of spontaneous EEG can track loss of consciousness across sleep and anaesthesia in humans [41–43]. Indeed, multivariate pattern analysis has shown that among all EEG features those that contribute most to the correct discrimination between conscious and unconscious patients are long-range connectivity and time-series complexity [44]. Further evidence that the dynamic complexity of intrinsic brain networks represents a reliable marker of consciousness has been provided by Demertzi and colleagues [45].

Given the sensitivity of network complexity measures in detecting a capacity for consciousness in challenging cases such as ketamine dreams and minimally conscious [39] and massively disconnected patients [11], they might be usefully applied to the even more challenging cases involving total disconnection. In principle, one could ask whether isolating, disconnecting, or growing brain islands allows for complexity levels comparable with the ones found in dreaming subjects, or in the injured brains of minimally conscious patients. Clearly, the thresholds and classifiers that currently allow accurate detection of consciousness in healthy subjects and neurological patients may become less reliable as we move toward atypical neural structures (Box 3). Nonetheless, these techniques would enable us to ask fundamental questions. Is it possible for *ex cranio* brains to display levels of complexity that are comparable with those displayed by the *in cranio* brains of conscious subjects? If so, for how long might that complexity be sustained? Could the disconnected hemisphere of a hemispherotomy patient host network dynamics that are as complex as the ones seen in the contralateral hemisphere, or does it plunge irreversibly into a deep sleep state due to lack of ascending

Please cite this article in press as: Bayne et al., Are There Islands of Awareness?, Trends in Neurosciences (2019), https://doi.org/10.1016/j.tins.2019.11.003

Box 4. Theories of Consciousness

One contrast between different theories of consciousness concerns the predictions that they make about when islands of awareness will (and will not) occur. Indeed, there is a more fundamental distinction between those theories of consciousness that allow for the possibility of islands of awareness and those that do not.

Those who are sympathetic to externalist accounts of consciousness are unlikely to hold that islands of awareness are possible, for externalists argue that the constitutive physical basis of consciousness extends beyond the brain and loops out into the body and environment [66,67]. If the existence of islands of awareness can be established, then that would place pressure on externalism.

Most theories of consciousness presuppose some version of internalism, and hold that the constitutive basis of consciousness is exclusively brain-bound. On this view, the body and the wider environment are only causally relevant to consciousness (see Box 2 for more on causality in this context). All internalist accounts hold that islands of awareness are possible, but different internalist theories make different predictions about the conditions under which islands of awareness might arise. For example, higher-order theories of consciousness predict that islands of awareness will arise when and only when a system for generating the appropriate kinds of higher-order representations is active [68]; global neuronal workspace accounts predict that islands of awareness require nonlinear global ignitions of activity involving (functional equivalents of) both parietal and frontal circuits [69]; and the Integrated Information theory predicts that brain islands will be conscious in a graded fashion depending on their intrinsic cause–effect power as measured by Φ [70]. In principle, then, the investigation of islands of awareness provides an important and informative constraint on theories of consciousness.

neuromodulation (Box 2)? Are there conditions in which cerebral organoids develop patterns of internal interactions that become progressively richer, or do such rich dynamics require a history of interaction with the environment? Addressing these questions would tell us much about the neural basis of consciousness.

Methods like PCI and other complexity measures may be able to detect an unsuspected capacity for consciousness in disconnected brains, but they cannot shed any light on the potential conscious contents that might occur in these islands. One in-principle approach would be to use stimulus-free neural decoding methods, such as those developed by [46] to decode the contents of dream experiences. Clearly, however, decoding conscious contents from within putative islands of awareness will be challenging. Although there is some evidence of robust mappings from neural activity to perceptual content in neurotypical subjects [47], it is doubtful that these mappings will also apply to atypical brains. Thus, the question of what kinds of contents might occur in islands of awareness is likely to present some of the deepest challenges hereabouts. We might be able to tell that there is something it is like to be a disconnected brain without being able to tell just what it is like.

Concluding Remarks

Suppose that we were to discover evidence of islands of awareness in *ex cranio* brains, hemispherotomy patients or cerebral organoids: what implications might such findings have (see Outstanding Questions)?

Let us begin with implications for accounts of consciousness (Box 4). The discovery of consciousness in a reanimated brain would indicate that ongoing dynamic interaction with the external world is not a necessary condition for consciousness. Although this finding would be in line with some accounts of consciousness, other accounts hold that neural systems are conscious only insofar as they are in ongoing dynamic interaction with their environment. Of course, even if a reanimated brain were able to sustain an island of awareness for a short period of time, it is entirely possible that external input (entrainment) would be needed for consciousness to be sustained in the longer term. The theoretical questions raised by cerebral organoids also concern the role that the environment plays in consciousness, although now the question is not only whether consciousness requires ongoing dynamic interaction with the environment, but whether it also requires a history of such interaction. Among the theoretical questions raised by hemispherotomy is whether consciousness requires subcortical input, or whether cortical activity alone might be sufficient to sustain consciousness. In short, the discovery of islands of awareness in any one of these three cases would reveal something important about the nature of consciousness, although different things would be revealed in each case.

A different set of implications that the discovery of islands of awareness would have is ethical and legal. The reanimation of *ex cranio* brains raises questions about personal identity and the nature of death. Suppose that whole brain reanimation were to occur in a human rather than a pig, and that pharmacological agents were not used to preclude organised neural activity. Would complex neural activity in such a brain, indicative of consciousness, ensure the continued existence of a particular individual, or does personal survival require (for example) a sense of bodily identity or agency? These questions are also raised by other conditions (such as dementia and serious brain injury), but they would be prompted in a particularly vivid fashion by the prospect of whole brain reanimation [48]. The discovery of an island of awareness in a hemispherotomy patient would also raise questions of personal identity. Would the experiences associated with the island of awareness belong to a separate subject of experience – a subject whose interests might diverge from those of the communicating subject – or should we think of the patient as a single subject of experience who happens to have two streams of awareness (one of which is isolated and one of which is not) [49,50].

The ethical challenges raised by the prospect of conscious organoids do not turn on questions of personal identity but on questions of moral status and standing [51–53]. How should conscious organoids be treated? What would it be to respect their well-being and interests? Should their treatment be governed by animal welfare laws, or would it require a new legal framework? Would it be permissible to engineer conscious organoids for research purposes, or would that violate their dignity? These questions should occupy a central place in the agenda of consciousness science.

In summary, advances in neurosurgery and neurotechnology may soon generate the capacity to create islands of awareness. It is not impossible that they may already have done so. It is imperative that the scientific and ethical consequences of these developments are subjected to careful consideration, alongside development of methods tuned to the detection of islands of awareness in the various different contexts in which they might arise.

Outstanding Questions

- Might (reversible) islands of awareness occur in contexts other than the ones that we have considered, such as epileptic absence seizures, or local islands of awareness in an otherwise sleeping brain?
- What other methods might there be for detecting islands of awareness?
- What animal models can be developed for investigating islands of awareness? For example, could animal hemispherotomy preparations be studied, with and without preservation of visual input to the otherwise disconnected hemisphere?
- What methods are there for interacting or communicating with an island of consciousness?
- Are islands of awareness possible for only short intervals, or could the disconnected brain sustain consciousness for prolonged periods of time?
- Are there some kinds of conscious contents that cannot occur in the disconnected brain, or that could only occur in such situations?
- Does the nature and distribution of islands of awareness discriminate between competing theories of consciousness?

Acknowledgments

T.B. is grateful to the Australian Research Council; A.K.S. is grateful to the Dr Mortimer and Theresa Sackler Foundation, which supports the Sackler Centre for Consciousness Science; M.M. is grateful to the European Union Horizon 2020 Framework Programme for its support through Grant Agreement No.785907 (Human Brain Project SGA2) and FETOPEN 2014-2015-RIA No. 686764 (Luminous); and all three authors are grateful for the support of the Canadian Institute for Advanced Research: Azrieli Program on Brain, Mind, and Consciousness. For helpful comments we thank Michael Carter and Adrian Owen.

References

1. Lagercrantz, H. and Changeux, J.P. (2009) The emergence of human consciousness: from fetal to neonatal life. *Pediatr. Res.* 65, 255–260
2. Menon, G.J. *et al.* (2003) Complex visual hallucinations in the visually impaired: the Charles Bonnet Syndrome. *Surv. Ophthalmol.* 48, 58–72
3. Griffiths, T.D. (2000) Musical hallucinosis in acquired deafness. Phenomenology and brain substrate. *Brain* 13, 2065–2076
4. Selimbeyoglu, A. and Parvizi, J. (2010) Electrical stimulation of the human brain: perceptual and behavioral phenomena reported in the old and new literature. *Front. Hum. Neurosci.* 4, 46
5. Marshel, J.H. (2019) Cortical layer–specific critical dynamics triggering perception. *Science.* Published online August 9, 2019. https://doi.org/10.1126/science.aaw5202
6. Callaghan, B.C. *et al.* (2009) Clinical reasoning: a 62-year-old woman with deafness, unilateral visual loss, and episodes of numbness. *Neurology* 72, e72–e78
7. Posner, J.B. *et al.* (2007) *Plum and Posner's Diagnosis of Stupor and Coma*, 4th ed, Oxford University Press
8. Smith, E. and Delargy, M. (2005) Locked-in syndrome. *BMJ* 330, 406–409
9. Nir, Y. and Tononi, G. (2010) Dreaming and the brain: from phenomenology to neurophysiology. *Trends Cogn. Sci.* 14, 88–100
10. Domino, E.F. (2010) Taming the ketamine tiger. *Anesthesiology* 113, 678–684
11. Rohaut, B. *et al.* (2017) Probing consciousness in a sensory-disconnected paralyzed patient. *Brain Inj.* 31, 1398–1403

12. Leonhard, S. *et al.* (2019) Diagnosis and management of Guillain-Barré syndrome in ten easy steps. *Nat. Rev. Neurol.* 15, 671–683
13. Friedman, Y. *et al.* (2003) Stimulation of brain death from fulminant de-efferentation. *Can. J. Neurol. Sci.* 30, 397–404
14. Joshi, M.C. *et al.* (2008) Guillain-Barré syndrome with absent brain stem reflexes – a report of two cases. *Anaesth. Intensive Care* 36, 867–869
15. Vargas, F. *et al.* (2000) Fulminant Guillain-Barré syndrome mimicking cerebral death: case report and literature review. *Intensive Care Med.* 26, 623–627
16. Cochen, V. *et al.* (2005) Vivid dreams, hallucinations, psychosis and REM sleep in Guillain-Barré syndrome. *Brain* 128, 2535–2545
17. Ragazzoni, A. *et al.* (2000) Event-related potentials in patients with total locked-in state due to fulminant Guillain-Barré syndrome. *Int. J. Psychophysiol.* 37, 99–109
18. Abdalmalak, A. *et al.* (2017) Single session communication with a locked-in patient by functional near-infrared spectroscopy. *Neurophotonics* 4, 040501
19. Forsberg, A. *et al.* (2008) Falling ill with Guillain-Barré syndrome: patients' experiences during the initial phase. *Scand. J. Caring Sci.* 22, 220–226
20. Liik, M. *et al.* (2012) Fulminant inflammatory neuropathy mimicking cerebral death. *BMJ Case Rep.* Published online July 20, 2012. https://doi.org/10.1136/bcr-10-2011-4906
21. Owen, A.M. *et al.* (2006) Detecting awareness in the vegetative state. *Science* 313, 1402
22. Monti, M.M. *et al.* (2010) Willful modulation of brain activity and communication in disorders of consciousness. *N. Engl. J. Med.* 362, 579–589
23. Vrselja *et al.*. (2019) Restoration of brain circulation and cellular functions hours post-mortem. *Nature* 568, 336–343
24. Mühlethaler, M. *et al.* (1993) The isolated and perfused brain of the Guinea-pig *in vitro*. *Eur. J. Neurosci.* 5, 915–926
25. de Curtis, M. and Llinas, R.R. (1993) Entorhinal cortex long-term potentiation evoked by theta-patterned stimulation of associative fibers in the isolated in vitro guinea pig brain. *Brain Res.* 600, 327–330
26. Ribaupierre, de S. and Delalande, O. (2008) Hemispherotomy and other disconnective techniques. *Neurosurg. Focus* 25 (E14), 1–10
27. Ibrahim, G.M. *et al.* (2015) Thalamocortical connectivity is enhanced following functional hemispherotomy for intractable lateralized epilepsy. *Epilepsy Behav.* 51, 281–285
28. Hawasli *et al.* (2017) Electrophysiological sequelae of hemispherotomy in ipsilateral human cortex. *Front. Hum. Neurosci.* Published online March 30, 2017. https://doi.org/10.3389/fnhum.2017.00149
29. Hertz-Pannier, L. *et al.* (2002) Late plasticity for language in a child's non-dominant hemisphere: a pre- and post-surgery fMRI study. *Brain* 25, 361–372
30. Timofeev, I. *et al.* (2000) Origin of slow cortical oscillations in deafferented cortical slabs. *Cereb. Cortex* 10, 1185–1199
31. Lancaster, M. *et al.* (2013) Cerebral organoids model human brain development and microcephaly. *Nature* 501, 373–379
32. Little, M.H. (2017) Organoids: a special issue. *Development* 144, 935–937
33. Sakaguchi, H. (2019) Self-organized synchronous calcium transients in a cultured human neural network derived from cerebral organoids. *Stem Cell Reports* 13, 458–473
34. Trujillo, C.A. *et al.* (2018) Nested oscillatory dynamics in cortical organoids model early human brain network development. *bioRxiv*. Published online September 3, 2018. https://doi.org/10.1101/358622
35. Naci, L. *et al.* (2014) A common neural code for similar conscious experiences in different individuals. *Proc. Natl. Acad. Sci. U S A* 111, 14277–14282
36. Faugeras, F. *et al.* (2011) Probing consciousness with event-related potentials in the vegetative state. *Neurology* 77, 264–268
37. Massimini, M. *et al.* (2009) A perturbational approach for evaluating the brain's capacity for consciousness. *Prog. Brain Res.* 177, 201–214
38. Casali, A.G. *et al.* (2013) A theoretically based index of consciousness. *Sci. Trans. Med.* 5, 198ra105
39. Sarasso, S. *et al.* (2015) Consciousness and complexity during unresponsiveness induced by propofol, xenon, and ketamine. *Curr. Biol.* 25, 3099–3105
40. Casarotto, S. *et al.* (2016) Stratification of unresponsive patients by an independently validated index of brain complexity. *Ann. Neurol.* 80, 718–729
41. Schartner, M.M. *et al.* (2015) Complexity of multi-dimensional spontaneous EEG decreases during propofol induced general anaesthesia. *PLoS One* 10, e0133532
42. Schartner, M.M. *et al.* (2017) Global and local complexity of intracranial EEG decreases during NREM sleep. *Neurosci. Conscious.* 2017, niw022
43. Li, D. and Mashour, G.A. (2019) Cortical dynamics during psychedelic and anesthetized states induced by ketamine. *Neuroimage* 196, 32–40
44. Sitt, J. *et al.* (2014) Large scale screening of neural signatures of consciousness in patients in a vegetative or minimally conscious state. *Brain* 137, 2258–2270
45. Demertzi, A. *et al.* (2019) Human consciousness is supported by dynamic complex patterns of brain signal coordination. *Sci. Adv.* 5, eaat760346
46. Horikawa, M. *et al.* (2013) Neural decoding of visual imagery during sleep. *Science* 340, 639–641
47. Nishimoto, S. *et al.* (2011) Reconstructing visual experiences from brain activity evoked by natural movies. *Curr. Biol.* 21, 1641–1646
48. Farahany, N. *et al.* (2019) Part-revived pig brains raise slew of ethical quandaries. *Nature* 568, 299–302
49. Bayne, T. (2010) *The Unity of Consciousness*, Oxford University Press
50. Schechter, E. (2018) *Self-Consciousness and "Split" Brains*, Oxford University Press
51. Hostiuc, S. *et al.* (2019) The moral status of cerebral organoids. *Regenerative Medicine* 10, 118–122
52. Lavazza, A. and Massimini, M. (2018) Cerebral organoids: ethical issues and consciousness assessment. *J. Med. Ethics* 44, 606–610
53. Shepherd, J. (2018) *Consciousness and Moral Status*, Routledge
54. Kondziella, D. *et al.* (2016) Preserved consciousness in vegetative and minimal conscious states: systematic review and meta-analysis. *J. Neurol. Neurosurg. Psychiatry* 87, 485–492
55. Gosseries, O. *et al.* (2014) Measuring consciousness in severely damaged brains. *Annu. Rev. Neurosci.* 37, 457–478
56. Plum, F. (1998) Coordinated expression in chronically unconscious persons. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 353, 1929–1933
57. Schiff, N. (1999) Words without mind. *J. Cognit. Neurosci.* 11, 650–656
58. Schiff, N. (2002) Residual cerebral activity and behavioural fragments can remain in the persistently vegetative state. *Brain* 125, 1210–1234
59. Block, N. (2002) The harder problem of consciousness. *J. Philosophy* 99, 391–425

60. Prinz, J. (2003) Level headed mysterianism. *J. Con. Studies* 10, 111–132
61. Phillips, I. (2018) The methodological puzzle of phenomenal consciousness. *Phil. Trans. R. Soc. Lond. B. Biol. Sci.* https://doi.org/10.1098/rstb.2017.034
62. Seth, A.K. *et al.* (2008) Measuring consciousness: relating behavioural and neurophysiological approaches. *Trends Cogn. Sci.* 12, 314–321
63. Shea, N. (2012) Methodological encounters with the phenomenal kind. *Philos. Phen. Res.* 84, 307–344
64. Shea, N. and Bayne, T. (2010) The vegetative state and the science of consciousness. *Br. J. Philos. Sci.* 61, 459–484
65. Chang, H. (2004) *Inventing Temperature*, Oxford University Press
66. Noë, A. (2006) Experience without the head. In *Perceptual Experience* (Gendler, T. and Hawthorne, J. eds), pp. 411–434, Oxford University Press
67. Thompson, E. and Cosmelli, D. (2011) Brain in a vat or body in a world? Brainbound versus enactive views of experience. *Phil. Top.* 39, 163–180
68. Brown, R. *et al.* (2019) Understanding the higher-order approach to consciousness. *Trends Cogn. Sci.* 23, 754–768
69. Deheane, S. and Changeux, J.P. (2011) Experimental and theoretical approaches to conscious processing. *Neuron* 70, 200–227
70. Tononi, G. and Koch, C. (2015) Consciousness: here, there and everywhere? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370, 20140167

Trends in Neurosciences

Cell Press
REVIEWS

Letter

Islands of Awareness or Cortical Complexity?

Benedetta Cecconi,[1,2]
Steven Laureys,[1,2] and
Jitka Annen[1,2,*]

Bayne *et al.* (2020) [1] envision the fascinating possibility that disconnected hemispheres (through hemispherotomy), *ex cranio* brains, and cerebral organoids may be islands of awareness (IOAs). They defined IOAs as systems fully disconnected from the external surrounding, both in terms of input (i.e., sensory information) and output (i.e., motor responses), yet capable of conscious experience (i.e., aware). In order to test whether these islands are effectively aware, the authors propose to use measures of human/animal consciousness that have been well validated in human/animal settings by correlation with pretheoretical measures of consciousness (e.g., behavioural reports) or subjective reports (in the case of humans). In particular, Bayne *et al.* [1] mention: (i) measures of dynamical complexity, capturing both integration and differentiation of the detected brain signal, such as the perturbational complexity index (PCI); and (ii) measures of algorithm complexity of EEG (i.e., entropy measures), quantifying the randomness of brain dynamics and capturing mainly the differentiation of the recorded brain activity. Two questions come to mind: one of a theoretical and one of a methodological kind. (i) Can subjective experience be collapsed to a 1D entity (i.e., a certain measure of cortical complexity)? (ii) If we accept such reduction, can measures validated in one context generalize to different species and neural systems?

If we accept that consciousness, now generally defined in the literature as subjective experience, can be assessed in isolated systems by the mere application of a certain measure of cortical complexity, consciousness is reduced to a 1D construct, in which an index is equated to (the capacity for) subjective experience. For example, let us consider the above-mentioned dynamic complexity (DC). DC operationalises a (phenomenological) structural property of consciousness, the integration and differentiation of every conscious scene (structural properties of consciousness are those properties common to all conscious experiences). While sufficiently high values of such measure appear necessary for consciousness, evidence that it is sufficient is lacking, as high DC could be present in unconscious networks [2]. In fact, consciousness would seem to possess other structural properties necessary for it to arise: for example, the features of perspectivalness (subjective first-person perspective), presentationality (to be in the present, the experience of 'nowness'), and transparency (attentional unavailability of earlier processing stages) [3]. Perhaps Bayne *et al.* [1] do not refer to subjective experience when speaking about consciousness in IOAs. In this case, it is legitimate to wonder what kind of consciousness would be identified if these systems were to be found dynamically complex: for example, will they possess a sense of phenomenal ownership, which seems necessary for the experience of suffering [3]? This is crucial for determining their moral status and regulating our behaviour toward them legally. Moreover, DC and entropy measures, quantifying different aspect of cortical complexity, may operationalise different phenomenological properties of conscious experience, potentially leading to conflicting interpretations of IOAs' state of consciousness. It is therefore necessary to specify which measure is to be considered to define an IOA as such.

Even if we accept the possibility of reducing consciousness to a certain measure of cortical complexity, we are still faced with the problem of defining a set of values consistent with consciousness. For example, electrically stimulated slices of the occipital cortex show a slice PCI score of 0.2 [4]; considered to index unconsciousness in the human PCI context [5]. Can we conclude that these slabs are unconscious? To help answer this question, let us imagine that human body temperature would be proposed as a universal index of health. A specific range of normal temperature has been set and validated as an index of human health by correlating it with human behaviours (i.e., sickness), such that if a woman at rest has a temperature of 40°C, we can conclude she is ill. We now apply this well-validated index to a bird. Since birds have an average body temperature of approximately 40°C, applying our index would make us conclude that it is sick. Similarly, the thresholds for (un)consciousness as defined by objective measures might not retain the same meaning when applied to other systems, without prior calibration. This calibration, requiring some form of subjective report/pretheoretical measure of consciousness, is by definition impossible to obtain in true islands of awareness. This raises the question whether consciousness can be quantified without relying on subjective/behavioural reports in systems very distant from us. Currently, human (or animal) consciousness is univocally probed through the subjective/behavioural reports of the person (or animal). Bayne and colleagues concluded that 'for the first time we might be able to tell that there is something it is like to be a disconnected brain without being able to tell just what it is like'[1], suggesting the presence of a subjective experience in IOAs. Or might we instead have reached the ultimate challenge of consciousness science, that is, to account for first-person phenomenology from a third-person perspective?

Acknowledgements

The authors thank the University and University Hospital of Liège, the GIGA Doctoral School for Health Sciences, the Belgian National Funds for Scientific Research (FRS-FNRS), the European Union's Horizon 2020 Framework Programme for Research and

Innovation under the Specific Grant Agreement No. 945539 (Human Brain Project SGA3), the European Space Agency (ESA), the Belgian Federal Science Policy Office (BELSPO) in the framework of PRODEX, the fund Generet and the King Baudouin Foundation. SL is research director at FRS-FNRS.

[1]Coma Science Group, GIGA-Consciousness, University of Liège, Liège, Belgium
[2]Centre du Cerveau, University Hospital of Liège, Liège, Belgium

*Correspondence:
jitka.annen@uliege.be (J. Annen).

https://doi.org/10.1016/j.tins.2020.05.007

References

1. Bayne, T. *et al.* (2020) Are there islands of awareness? *Trends Neurosci.* 43, 6–16
2. Seth, A.K. *et al.* (2006) Theories and measures of consciousness: an extended framework. *Proc. Natl. Acad. Sci. U. S. A.* 103, 10799–10804
3. Metzinger, T. (2003) *Being No One : the Self-Model Theory of Subjectivity*, MIT Press
4. D'Andola, M. *et al.* (2018) Bistability, causality, and complexity in cortical networks: an *in vitro* perturbational study. *Cereb. Cortex* 28, 2233–2242
5. Casali, A.G. *et al.* (2013) A theoretically based index of consciousness independent of sensory processing and behavior. *Sci. Transl. Med.* 5, 198ra105

Trends in Neurosciences

CellPress REVIEWS

Letter

From Complexity to Consciousness

Tim Bayne,[1,2,*]
Anil K. Seth,[2,3,4] and
Marcello Massimini[2,5,6]

We thank Cecconi and colleagues for their probing questions about our opinion article 'Are There Islands of Awareness?' [1], and we are grateful for the opportunity to return to these important issues.

Our primary aim in writing the aforementioned paper was to give the notion of an island of awareness (IOA) some precision, and to argue that the possibility of IOAs should not be dismissed as a philosophical fantasy but should be taken seriously by the science of consciousness.

The fundamental challenge, concerns the detection of IOAs. What would count as good evidence for the presence of an IOA? We suggest that the most promising approaches involve complexity measures, such as the Perturbational Complexity Index (PCI), and other measures, which are sensitive to the coexistence of functional differentiation and functional integration in neural systems. We point to these particular metrics because they go beyond mere correlation to operationalize explanatory relationships between neural dynamics and phenomenological properties [2,3], and because they have already been shown to perform well when assessing consciousness in patients in whom devastating injuries often result in disconnection from inputs and outputs [4,5]. Thus, we propose to use a 'radar' that we have learned to trust to explore the unknown landscape of completely disconnected brains.

Cecconi *et al.* [6] raise two interesting issues in connection with this proposal. The first is that these measures treat consciousness as a one-dimensional phenomenon, which, Cecconi and colleagues argue, is mistaken. We agree that consciousness is multidimensional [7], but we don't agree that the use of complexity measures to detect consciousness presupposes that it is unidimensional. The key here is to realize that markers or indicators of consciousness need not be treated as definitions of consciousness. (Compare: crying is a marker of pain, but pain is not to be defined in terms of crying.) Consciousness could be multidimensional even if certain markers of it are unidimensional. (Cecconi and colleagues ask whether we mean to refer to subjective experience when speaking of consciousness. We do.)

The second issue that Cecconi *et al.* raise concerns the use of complexity measures to identify IOAs. There are two aspects to this challenge. The first is whether it is possible to directly compare values of complexity that have been derived from different types of systems. For example, one cannot directly compare the values of the slice-adapted version of PCI (sPCI) to those validated in humans, in part because they are currently obtained with different techniques at different scales [8]. Determining what counts as the 'most appropriate' scale for a given system is challenging, but recent advances in the mathematical analysis of multiscale systems have made progress in addressing it [9,10].

Accompanying these technical considerations is the more general question of whether it is ever legitimate to apply markers of consciousness that have been validated in one population (intact human beings) to the members of other populations, such as ex-cranio brains, disconnected hemispheres, and cerebral organoids. Cecconi *et al.* appear to answer this question in the negative. We will confine our response to three points. First, anyone who is willing to ascribe consciousness to non-human animals faces a version of this challenge, for the behavioural and neurophysiological markers of consciousness that we employ to ascribe consciousness to non-human animals must themselves be validated with reference to human beings [11]. Second, the systems we are considering share many of the neurophysiological features of the brains of intact mammals; in particular, they have neurons and associated biological infrastructure. Concerns about the possible substrate dependency of consciousness can therefore be set aside [compare, for example, silicon-based artificial intelligence (AI) systems]. Third, as we note above and in [1], our validation strategy appeals to the capacity of the property in question to explain features that are associated with consciousness, starting from its unity and richness. This approach to validation (i.e., the 'natural kind' [12] or 'explanatory correlate' [13] approach) arguably supports greater confidence in the applicability of measures of consciousness to 'novel populations' than do correlational approaches to validation. For this reason, the analogy that Cecconi *et al.* draw with temperature and health fails.

There is always some risk in applying a marker of consciousness to 'difficult cases'. The central issue is one of risk management: would the level of risk associated with ascribing consciousness to disconnected hemispheres, ex-cranio brains, or cerebral organoids be unreasonable? We take a pragmatic view. The recent emergence of candidate IOAs in clinical settings and in neurotechnology laboratories raise questions about consciousness that cannot be ignored. Although current complexity measures are imperfect and must be interpreted in context, they remain our most promising guide to this uncharted terrain.

[1]School of Philosophical, Historical, and International Studies, Monash University, Melbourne, Australia
[2]Canadian Institute for Advanced Research, Azrieli Program in Brain, Mind, and Consciousness, Toronto, Canada

[3]School of Engineering and Informatics, University of Sussex, Brighton, UK
[4]Sackler Centre for Consciousness Science, University of Sussex, Brighton, UK
[5]Department of Biomedical Clinical Sciences, University of Milan, Milan, Italy
[6]IRCCS, Fondazione Don Gnocchi, Milan, Italy

*Correspondence:
timothy.bayne@monash.edu (T. Bayne).

https://doi.org/10.1016/j.tins.2020.05.008

References

1. Bayne, T. *et al.* (2020) Are there islands of awareness? *Trends Neurosci.* 43, 6–16
2. Seth, A.K. *et al.* (2011) Causal density and integrated information as measures of conscious level. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* 369, 3748–3767
3. Tononi, G. and Edelman, G.M. (1998) Consciousness and complexity. *Science* 282, 1846–1851
4. Casarotto, S. *et al.* (2016) Stratification of unresponsive patients by an independently validated index of brain complexity. *Ann. Neurol.* 80, 718–729
5. Demertzi, A. *et al.* (2019) Human consciousness is supported by dynamic complex patterns of brain signal coordination. *Sci. Adv.* 5, eaat7603
6. Cecconi B, *et al*. Islands of awareness or cortical complexity? *Trends Neurosci.* (in press). https://doi.org/10.1016/j.tins.2020.05.007
7. Bayne, T. *et al.* (2016) Are there levels of consciousness? *Trends Cogn. Sci.* 20, 405–413
8. D'Andola, M. *et al.* (2018) Bistability, causality, and complexity in cortical networks: an *in vitro* perturbational study. *Cereb. Cortex* 28, 2233–2242
9. Tononi, G. *et al.* (2016) Integrated information theory: from consciousness to its physical substrate. *Nat. Rev. Neurosci.* 17, 450–461
10. Rosas, F. *et al.* (2020) Reconciling emergences: an information-theoretic approach to identify causal emergence in multivariate data. *ArXiv* Published online April 17 2020. https://arxiv.org/abs/2004.08220
11. Edelman, D.B. and Seth, A.K. (2009) Animal consciousness: a synthetic approach. *Trends Neurosci.* 32, 476–484
12. Shea, N. and Bayne, T. (2010) The vegetative state and the science of consciousness. *Br. J. Philos. Sci.* 61, 459–484
13. Seth, A.K. (2009) Explanatory correlates of consciousness: theoretical and computational challenges. *Cogn. Comput.* 1, 50–63

Review

Animal consciousness: a synthetic approach

David B. Edelman[1] and Anil K. Seth[2]

[1] The Neurosciences Institute, 10640 John Jay Hopkins Drive, San Diego, CA 92121, USA
[2] Department of Informatics, University of Sussex, Brighton BN1 9QJ, UK

Despite anecdotal evidence suggesting conscious states in a variety of non-human animals, no systematic neuroscientific investigation of animal consciousness has yet been undertaken. We set forth a framework for such an investigation that incorporates integration of data from neuroanatomy, neurophysiology, and behavioral studies, uses evidence from humans as a benchmark, and recognizes the critical role of explicit verbal report of conscious experiences in human studies. We illustrate our framework with reference to two subphyla: one relatively near to mammals – birds – and one quite far –cephalopod molluscs. Consistent with the possibility of conscious states, both subphyla exhibit complex behavior and possess sophisticated nervous systems. Their further investigation may reveal common phyletic conditions and neural substrates underlying the emergence of animal consciousness.

A synthetic framework for studying animal consciousness

Although Darwin proposed that animal and human minds alike are the products of natural selection [1], questions of animal consciousness were largely neglected throughout the 20th century (but see Griffin [2]; note the term 'animal' is used here to mean 'non-human animal'). This neglect may have arisen in part because, seemingly, only humans are capable of accurately describing their phenomenal experience. However, there is now abundant and increasing behavioral and neurophysiological evidence consistent with, and even suggestive of, conscious states in some animals. We will use humans as a benchmark for the development of new empirical criteria for further investigation. Here we apply this approach to birds and cephalopod molluscs, subphyla that exhibit complex cognitive faculties and behaviors and have strikingly elaborate brains. These two subphyla are examples of highly distinct lineages, and their study provides an excellent opportunity to examine how conscious states might be instantiated in very different nervous systems. While we do not resolve this issue here, we propose that its examination lies within the reach of contemporary neuroscience.

Humans as a benchmark

The notion that consciousness can be engendered in different nervous systems by a variety of underlying mechanisms suggests a need to examine constraints, and therefore to synthesize behavioral, neurophysiological, and neuroanatomical evidence. Human studies involving the correlation of *accurate report* with neural correlates can provide a valuable benchmark for assessing evidence from studies of animal behavior and neurophysiology. A constraint on this strategy is that the capacity for accurate report of conscious contents implies the presence of *higher-order consciousness*, which in advanced forms may require linguistically-based narrative capability. This is in contrast to *primary consciousness*, which entails the ability to create a scene in the 'remembered present' [3] in the absence of language. Primary consciousness may be a basic biological process in both humans and animals lacking true language.

Various properties of human consciousness can be identified at the neural, behavioral, and phenomenal levels [4]. Neural correlates of human consciousness include the presence of thalamocortical signaling, fast, irregular,

Corresponding author: Edelman, D.B. (david_edelman@nsi.edu).

Glossary

Accurate report: a first-person account of what an individual is experiencing, made without the attempt to mislead. Accurate report, which can be given through language or related varieties of voluntary response, has been critical in the investigation of conscious states in humans. In animals without the faculty of natural language, forms of behavioral report acting through other motor channels might be examined to determine the possible presence of high-order discriminations suggesting conscious states.

Binocular rivalry: this occurs when the two eyes are each simultaneously presented with a different image. Rather than seeing both images superimposed on one another, the subject sees one image first, then the other, in an alternating sequence. For example, if one eye is presented with parallel vertical stripes and the other with parallel horizontal stripes, rather than seeing an overlapping 'weave' of vertical and horizontal stripes, the subject sees first one orientation of stripes, then the other [83].

Explanatory correlates of consciousness: Conventional approaches within consciousness science have emphasized the search for so-called 'neural correlates of consciousness': neural activity having privileged status in the generation of conscious experience [7]. However, the transition from correlation to explanation requires an explanation of how particular neural correlates *account for* specific properties of consciousness. Searches for explanatory correlates of consciousness attempt to provide this link [11,14].

Primary consciousness: this refers to the experience of a multimodal scene composed of basic perceptual and motor events. Primary consciousness is sometimes called perceptual or phenomenal consciousness, and it may be present in animals without true language.

Higher-order consciousness: this, by contrast, involves the referral of the contents of primary consciousness to interpretative semantics, including a sense of self and, in more advanced forms, the ability to explicitly construct past and future narratives [3]. The presence of higher order consciousness or metacognition should not be assumed to be necessary for the ascription of primary consciousness, though it may be constitutively required for advanced forms of self-consciousness and consciousness of consciousness (Box 2).

Transitive inference: the ability to connect two or more separate relations, and is widely regarded as a fundamental process in reasoning. Such abilities have been demonstrated in certain birds. For example, in experiments involving the relative ranking of objects presented in pairs, pigeons were able to determine that B>D after they had separately learned that A>B, B>C, C>D, and D>E, while great tits have been observed to deduce complex social dominance rankings [30].

0166-2236/$ – see front matter doi:10.1016/j.tins.2009.05.008 Available online xxxxxx

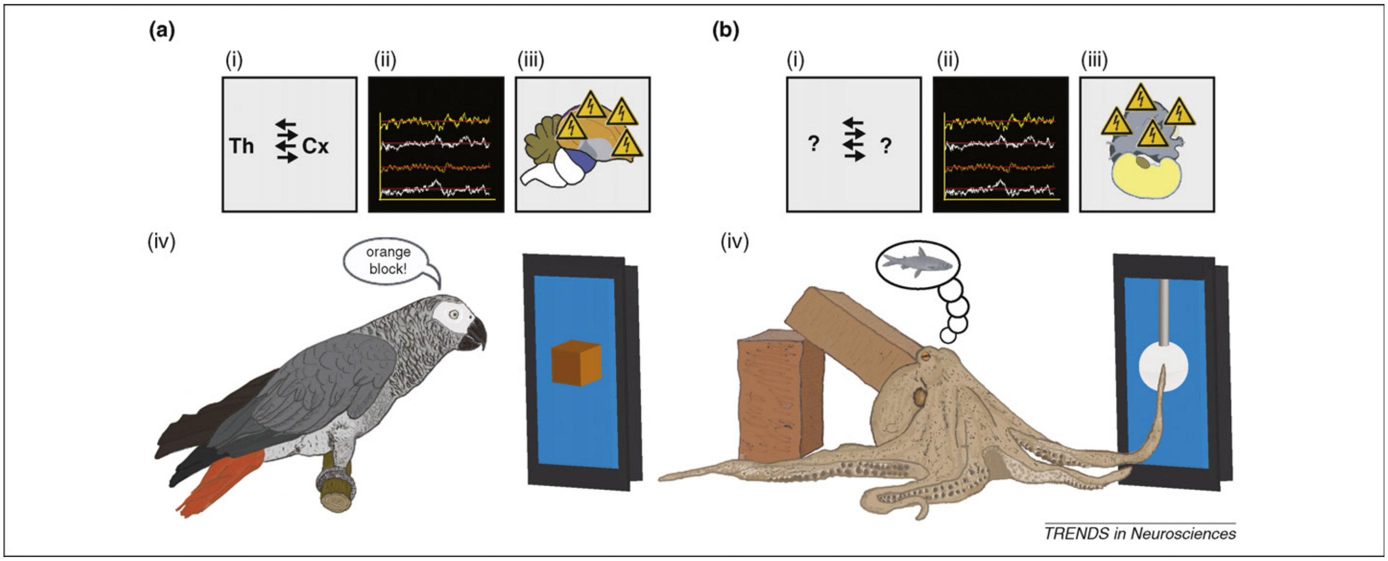


Figure 1. The investigation of possible conscious states in non-human species as disparate as birds **(a)** and cephalopods **(b)** can be informed by searching for neural properties that have been correlated with consciousness in humans, including reentrant signaling between thalamus and cortex **(a, i)** or putative functional analogs **(b, i)**, fast, irregular, low-amplitude EEG signals (**a** and **b, ii**), and widespread electrical activity in cortex **(a, iii)** or functionally analogous structures **(b, iii)**. Such processes in animals can best be related to consciousness when they can be correlated with accurate reports. Relevant forms of report include vocalizations in the case of African grey parrots (**a, iv**) or coloration and body patterning in the case of cephalopods **(b, iv)**. In the figure, an African grey parrot and a common octopus (*O. vulgaris*) respond to salient artificial stimuli presented on video displays: an orange block in a discrimination task **(a, iv)** and a white ball that has been previously associated with food (herring) during training **(b, iv)** (see Box 1).

low-amplitude electroencephalographic (EEG) signals, and widespread cortical activity correlated with conscious contents [5–7]. At the behavioral level, consciousness has been associated with behavioral flexibility [8], rational action [9], and certain forms of conditioning [10]. These can be related to cognitive properties involving widespread access and associativity [8], multiple discriminations [11], and the capacity for accurate report [5]. These properties can be mapped to a variety of functions related to consciousness [12]. At the phenomenal level, human consciousness involves the presence of a sensorimotor scene, the existence of a first-person perspective, the experience of emotions, moods, and a sense of agency [13–14].

Using humans as a benchmark, behavioral, cognitive, and neural properties can be employed as empirical criteria informing the ascription of conscious states to animals (Figure 1). The application of this approach requires that: (i) at least some of these properties reliably occur in both humans and animals; (ii) human brain areas responsible for consciousness can be seen to be integrated with those areas responsible for accurate report of phenomenal experience; and (iii) neural evidence that can be correlated with phenomenal properties of consciousness must in addition *account for* those properties.

In humans, explicit verbal, or linguistic, report of a conscious experience is sometimes taken as a 'gold standard' in the sense that it guarantees the presence of consciousness. However, many creatures, including infants, most animals, and aphasic human adults, are constitutively unable to produce linguistic reports. The production of such reports is therefore too limiting a criterion for the ascription of consciousness in general. Importantly, accurate report may exploit behavioral channels other than language, for example lever presses or eye blinks (Box 1). However the ability to provide such report nonetheless implies the presence of higher-order (meta-cognitive) access to primary conscious contents, which may not be constitutively required for primary consciousness. Our approach therefore recognizes that the mechanisms responsible for primary consciousness may be distinct from those mechanisms enabling its report.

The extent to which neural evidence can account for phenomenal properties is particularly important with respect to those properties that are common to most or all conscious experiences. For example, in humans, every conscious scene is both integrated (i.e. 'all of a piece') and differentiated (i.e. composed of many different parts) [11]. Therefore, finding neural processes that themselves exhibit simultaneous integration and differentiation would help to *explain*, and not merely correlate with, the corresponding phenomenal property. Such neural processes can therefore be considered to be '*explanatory correlates of consciousness*' [14], and because they are explanatory, their identification in animals is more suggestive of the presence of corresponding phenomenal properties than is the identification of neural correlates *per se*. In this view, conscious states are neither identical to neural states nor are they computational or functional accompaniments to such states; rather, conscious states are entailed by neural states in much the same way that the spectroscopic response of hemoglobin is entailed by its molecular structure [6].

The presence of voluntary behavioral responses is another candidate 'gold standard' for the ascription of consciousness. However, the absence of such responses in humans does not necessarily imply absence of phenomenal experience. For example, in a recent study of a patient in a behaviorally unresponsive vegetative state, brain activity related to volition was taken as persuasive evidence of residual consciousness [15]. Conversely, behavior that appears to be volitional could be attributed widely among animals on the basis of spontaneous and adaptive

Box 1. Metacognition, blindsight, and behavioral report

In seminal experiments by Cowey and Stoerig [84], rhesus macaques with lateralized lesions to the visual area V1 were trained to touch a region of a video display where stimuli appeared. These monkeys could detect and discriminate between stimuli presented within their lesioned hemifields, but they could not distinguish between regions with stimuli presented in these hemifields and regions containing no stimuli presented in their unaffected hemifields. The absence of response in the latter case has been interpreted as a 'metacognitive comment', indicating lack of awareness of the difference between the two hemifields. This could be compared to the responses of human 'blindsight' patients who claim not to see stimuli to which they nonetheless successfully respond. The plausibility of this comparison depends to a degree on the highly conserved neuroanatomy between macaques and humans. When adapted in conjunction with an improved understanding of neural analogs and homologs between mammals and other subphyla, the 'commentary key' paradigm suggested by Cowey and Stoerig [84] provides a valuable experimental platform. For example, 'avian blindsight' might be inducible by lesioning the ectostriatum, the avian brain area analogous to mammalian V1 (Figure 2). Although the optic tracts of many birds are nearly completely decussated (>99%) to opposite hemispheres of the brain, precluding hemifield arrangements analogous to those in monkeys and humans, and the avian brain lacks a corpus collosum, evidence of interhemispheric switching [85] suggests that binocular rivalry might be achieved under appropriate experimental conditions. Metacognitive comments on the presence or absence of stimuli in occluded hemifields could be made through vocalizations previously entrained to those stimuli.

Similar approaches can be envisaged in cephalopod molluscs if functional analogs to mammalian visual cortex can be discerned (Figure 1). For example, we suggest a version of the 'attentional blink' paradigm, a phenomenon observed when human subjects are presented with a rapid sequence of co-located visual stimuli. A subject will fail to observe a secondary salient target stimulus occurring within this succession if it is presented between 200 and 500 ms after the first stimulus [83]. In an octopus presented with a serial stream of stimuli containing a 'blink' stimulus, the chromatophore system [86,87] or other components of body patterning [88] might provide a channel for experiments to explore similar effects. For example, an octopus could be presented with a serial stream containing an image of an octopus displaying chromatophore patterns associated with aggression in succession with patterns associated with mating display. Variation in responses to the second target, if presented in close temporal proximity to the first, would be suggestive of a 'blink.' Such 'psychophysical' experiments must recognize that cephalopod vision is substantially different from vertebrate vision. Most cephalopod molluscs are color-blind [89], but many are known to perceive polarized light and make subtle discriminations based on this capacity [90,91].

behavioral responses. Therefore, apparent voluntary behavior at best provides a weak criterion for the ascription of consciousness. In addition, conditions such as the vegetative state underscore the point that consciousness should not be confused with arousal or inferred directly from the existence of distinct sleep/wake cycles. States resembling deep sleep have been observed in many animals, including *Drosophila melanogaster* [16] and *Caenorhabditis elegans* [17]; conversely, female killer whales and dolphins and their newborn calves may not sleep for periods of four to six weeks *postpartum* [18].

Mammalian consciousness: extending the benchmark

Mammals, particularly primates, share with humans many neurophysiological and behavioral characteristics relevant to consciousness, and therefore represent a relatively uncontroversial case for the ascription of at least primary consciousness [5]. In a classical example, Logothetis *et al.* trained rhesus macaque monkeys to press a lever to report perceived stimuli in a *binocular rivalry* paradigm [19] (Box 1). Neurons in macaque inferior temporal (IT) cortex showed activity correlated with the reported percept, whereas neurons in the visual area V1 instead responded to the visual signal. This suggests a critical role for IT in visual consciousness. These observations are consistent with evidence from humans subjected to binocular rivalry while being examined via magnetoencephalography (MEG). The results from these studies suggest that consciousness of an object involves widespread coherent synchronous cortical activity [20]. This correspondence between monkeys and humans provides an example of how benchmark comparisons across humans and animal species can be made. With this in mind, we now explore the application of benchmark comparisons to two widely divergent animal subphyla: birds and cephalopod molluscs. We are aware of the pitfalls of making facile comparisons and implying that homologies exist in the absence of strong evolutionary evidence. In each case, we will present behavioral evidence first, and then present evidence based on neural architecture and dynamics.

Building a case for avian consciousness

Avian cognition and behavioral capabilities

Feats of avian memory [21,22], tool use and manufacture [23], deception [24], and vocal learning and performance [25,26], the capacity of some species to employ lexical terms in meaningful ways [27], and evidence for higher order discriminations in some birds [27,28], collectively support the functioning of nervous systems as sophisticated as those of most mammals.

Episodic and working memory capabilities are implied by the sophisticated food caching behaviors of corvids (e.g. jays, jackdaws, magpies, rooks, crows, and ravens) [23] and by laboratory-based demonstrations of *transitive inference* and delayed-match-to-sample in pigeons and great tits [29]. African grey parrots, magpies, and ring doves have shown the ability to track periodically hidden and displaced objects; such object constancy certainly requires working memory [30]. In addition, spatial learning, though obviously implied by navigation during flight, has been shown explicitly by hooded crows learning to negotiate a radial maze [31].

Studies of avian tool use and manufacture imply not only elaborate memory and learning substrates, but also the ability to make sophisticated discriminations and to plan behaviors before executing them. For example, New Caledonian crows have been observed to fashion hooked-wire tools to retrieve food from a glass cylinder, sometimes flying to a distant perch to bend the wire before returning to the cylinder [23]. Similarly, wild crows are known to fabricate tools from twigs and leaves in order to extract insects from holes in trees [23]. In addition, a variety of avian species, including Japanese quails [32] and European starlings [33], may be capable of social or observational learning (but see Zentall [34] and Heyes [35]). Perhaps most notable are instances in which scrub jays re-cached food in private after their initial caching was witnessed by conspecifics [30], and observations of ravens challenging conspecifics that witnessed caching activities while ignoring 'naïve' birds [36]. These behaviors suggest that some birds may be capable of theory of mind.

The avian vocal channel

While the foregoing cognitive capabilities are suggestive of conscious states, the most promising avenue for investigating avian consciousness may involve the study of species capable of vocal learning, which enables a highly flexible form of accurate report. The capacity for some form of vocal learning is shared by at least six animal groups, including cetaceans, bats, parrots, songbirds, hummingbirds, elephants, and possibly even mice and some other rodents [37]. In birds, vocal learning enables sophisticated song learning and production, mimicry of sounds, and, in the psittacines (parrots), word production, comprehension, and naming [27]. For example, African grey parrots were able to name objects, having acquired vocabularies roughly equivalent to those of some language-trained chimpanzees (albeit after years of training and reinforcement) [38]. Indeed, by naming objects in categorization paradigms, these animals appeared to produce accurate reports of sophisticated discriminations they were making. 'Alex', a principal subject of many of these experiments, when presented with an altered array of objects, seemed able to make a judgment to the effect that – 'I know that something in this perceptual scene has changed, and here is what has changed'. This finding suggests an ability to make discriminations about putative primary conscious states that appears to resemble some form of higher order consciousness [27,28].

Avian neural structures and processes homologous to those of mammals suggest possible neural substrates for both consciousness and its report

On what structural bases might avian vocal behavior be related to structures underlying human verbal accurate report? Supporting evidence could come from shared neural mechanisms. One example may be the neural substrate for motor learning in mammals and that for song learning in some birds. Much of the neural basis for song learning in oscines (songbirds) and psittacines resides in an anterior forebrain pathway involving the basal ganglia, in particular, a striatal nucleus called Area X [39] (see Figure 2). The anatomical and physiological properties of neurons in Area X closely resemble those of neurons in the mammalian striatum. Specifically, the four neuronal phenotypes found in mammalian striatum are also present in Area X. A notable difference is the presence of a fifth neural phenotype in Area X – but not in mammalian striatum – that is similar to cells found in the mammalian globus pallidus. Area X may therefore comprise a novel mixture of striatal and pallidal anatomies, but it is nonetheless recognizably homologous to the direct striatopallidothalamic pathway of the mammalian basal ganglia [39]. A similar circuit has been reported in the anterior forebrain of the budgerigar, a parrot [40]. Together, these findings strongly suggest common functional circuitry underlying the organization and sequencing of motor behaviors related to vocalization in birds and mammals capable of vocal learning. However, whether in some birds this circuitry is embedded within a broader network homologous to that underlying human verbal report remains to be determined.

Another avenue for exploring non-mammalian consciousness is to identify structural and functional homologs to mammalian thalamocortical systems. Vertebrate nervous systems follow a highly conserved body plan that emerged with the first chordates more than 500 million years ago. Consequently, many vertebrate neural structures can be traced to common origins in specific embryological tissues. Avian homologs of subcortical structures, such as the hypothalamus and pre-optic area, are relatively easy to recognize. Although the identity of the avian neural homolog of mammalian isocortex remains controversial [41,42], comparative embryological studies suggest that the basic underlying neuronal composition and circuitry of the mammalian cortex were established within clustered arrangements of nuclei long before the appearance of the distinct six-layered mammalian cortex [43]. In particular, the nuclei comprising the dorsal ventricular ridge (DVR) of the developing avian brain contain neuronal populations homologous to those present in different layers of the mammalian neocortex. These include neurons receiving thalamic input, as well as cells projecting to brainstem and spinal cord neurons. The neurons of the avian DVR and mammalian cortex are nearly identical in both their morphology and constituent physiological properties [44].

Structural homologies can also be identified using molecular and immunohistological techniques. In particular, neurotransmitters, neuropeptides, and receptors specific to particular neuronal populations within mammalian brain regions have been localized to homologous avian brain regions. For example, both AMPA receptor subunits and the pallidal neuron/striatal interneuron marker Lys8-Asn9-neurotensin8–13 (LANT6) are found in the neurons of both mammalian and avian basal ganglia [45,46]. Finally, gene expression patterns similar to those of mammals have been identified in the avian brain. For example, a comparison of homeotic genes involved in early brain development in chick and mouse embryos has revealed robust structural homologies between parts of the avian telencephalon and mammalian cortex [47].

Deep avian–mammalian homologies have also been revealed by examining functional properties of neuronal populations within particular brain regions. The avian anterior forebrain pathway may be functionally analogous to the mammalian cortico-basal ganglia-thalamocortical. This is suggested by the presence of both inhibitory and excitatory pathways in the medial nucleus of the avian dorsolateral thalamus (DLM), as well as by functional similarities between neurons in the DLM and mammalian thalamocortical neurons. Similarities have also been found among the excitatory and inhibitory circuitry of birds and mammals, particularly in the serotonergic, GABAergic, and dopaminergic systems [48].

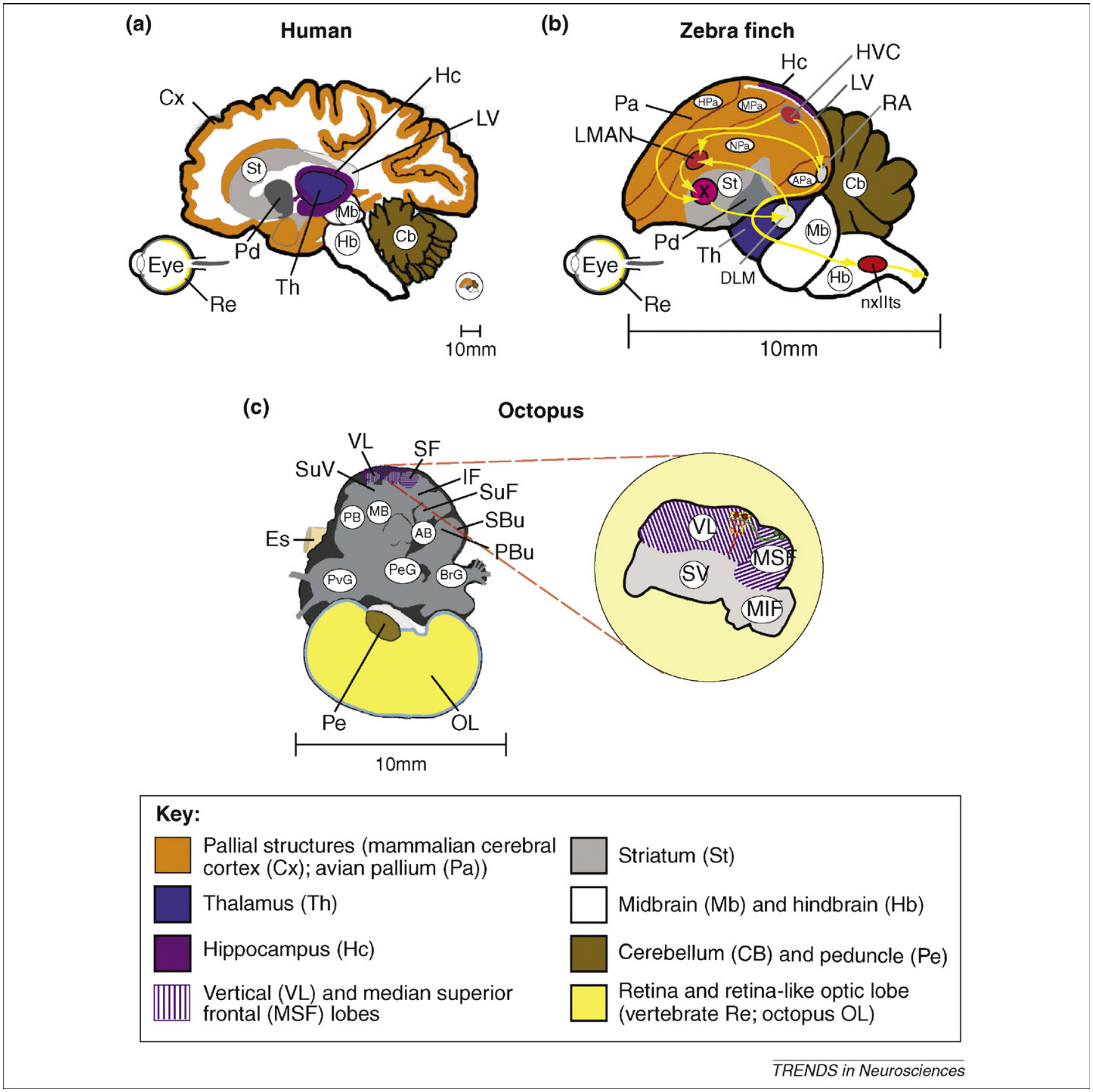


Figure 2. Avian and mammalian brains contain homologous structures and similar functional circuitry; the complex nervous systems of coleoid cephalopods may exhibit some functional circuitry analogous to those of higher vertebrates. **(a)** Midline sagittal section of a human brain showing major structures, including those involved in generating conscious states (e.g. cortex, thalamus, and basal ganglia). **(b)** Midline sagittal section of the brain of a zebra finch, a songbird. Major neural structures are shown, including those with mammalian homologs. Also shown is a greatly simplified schematic of the anterior forebrain pathway for song learning (yellow arrows) involving components of the basal ganglia, including the striatal nucleus Area X ('X' in filled red circle). The circular inset to right of human brain shows zebra finch brain to scale for comparison. **(c)** Midline section of the brain of an octopus (*O. vulgaris*). Most major lobes and ganglia are shown. Vertical (VL) and medial superior frontal (MSF) lobes (purple hatched lines), containing circuitry critical for long-term memory, are shown in a magnified view in circular inset on right. The key (bottom of figure) shows color-coding of major brain regions to indicate homology or functional and/or structural analogy. Other regions of human and avian brains are labeled: LV, lateral ventricle; Pd, pallidum. Pallial divisions of the avian cerebrum are indicated as follows: HPa, hyperpallium; MPa, mesopallium; NPa, nidopallium. Components of the avian anterior forebrain pathway are indicated as follows: DLM, medial nucleus of the dorsolateral thalamus; HVC, higher vocal center; LMAN, lateral magnocellular nucleus of the anterior neostriatum; nxIIts, tracheosyringeal portion of hypoglossal nucleus; RA, robust nucleus of the archistriatum. Major lobes and ganglia of the octopus brain are indicated as follows: AB, anterior basal lobe; BrG, brachial ganglia; Es, esophagus; IF, inferior frontal lobe; MB, median basal lobe; MIF, medial inferior frontal lobe; PBu, posterior buccal lobe; Pe, peduncle; PeG, pedal ganglia; PB, posterior basal lobe; PvG, palliovisceral ganglia; SuV, subvertical lobe; SBu, superior buccal lobe; SF, superior frontal lobe; SuF, subfrontal lobe. Scale bars are shown at the bottom of each brain section.

In addition to neuroanatomy, electrophysiological studies are critical in establishing functional homologies between avian and mammalian nervous systems. Currently, however, common properties of mammalian thalamocortical neurons, such as low-threshold calcium (Ca^{2+}) spikes and slow oscillations, have not yet been found in birds [49]. Nonetheless, similarities between the waking EEG patterns of birds and mammals, as well as slow wave electrical activity recorded during avian sleep [50], are suggestive of neural dynamics that might support conscious states in birds.

The existence in birds of structural and functional homologies to mammalian thalamocortical systems is certainly consistent with the presence of higher cognitive faculties and perhaps consciousness [30,51]. Nevertheless, a compelling case for avian consciousness cannot be made solely on the strength of relevant neuroanatomical and neurophysiological resemblances. Nor are descriptions of avian behaviors that imply sophisticated cognitive capabilities sufficient to make such a case. New experimental strategies are needed for the evaluation of possible conscious discriminations in awake, behaving birds. The findings obtained from studies of Alex the African grey parrot encourage the development of such strategies. Even more challenging though are approaches to investigating conscious behavior in invertebrates.

Searching for consciouness in invertebrates: The cephalopod case

The richness of cephalopod behavioral repertoires

In addition to possessing large brains, cephalopod molluscs have extremely flexible behavior and highly developed attentional and memory capacities that may be suggestive of conscious states [52]. The performance of some cephalopods (particularly *Octopus vulgaris*) in several learning and memory paradigms (e.g. flexibility, persistence of memory traces, contextual learning) [53] is formidable, and comparable in sophistication to that of some vertebrates. Octopuses can make discriminations between different objects based on size, shape, and intensity [54,55], classifying differently shaped objects in the same manner as vertebrates ranging from goldfish to rats [55]. Octopuses are also capable of finding the correct path to a reward in a plexiglas maze and can retrieve objects from a clear bottle sealed with a plug [56,57]. In another striking study, 'naïve,' or 'observer,' octopuses watched conditioned animals ('demonstrators') choose between two simultaneously presented objects that differed in contrast only; the observer octopuses later made the same contrast choices in isolation and without any explicit conditioning [58]. Although controversial [59–62], this finding suggests that octopuses are capable of observational learning, a faculty previously thought to be unique to highly social animals.

Finally, distinct capacities for short- and long-term memory have been shown in both the octopus and the cuttlesh [53,63]. In a maze containing obstacles that were changed *ad libitum*, octopuses could remember these changes and adjust their movements accordingly. Interestingly, the octopuses in this study appeared to pause and deliberate about the layout of the maze before proceeding [64].

Cephalopod brains: complex nervous systems distant from the vertebrate line

The organization of invertebrate nervous systems diverges so greatly from those of vertebrates such as birds and mammals that, until recently, sophisticated cognitive capabilities had rarely been ascribed to invertebrate species (see Figure 2). Bees [65,66], spiders [67,68], and the cephalopod molluscs [52] are notable exceptions.

Of all cephalopod molluscs, the octopus has the largest population of sensory receptors. These receptors communicate with a nervous system that, in adults, may contain between 170 million and 500 million cells, most of which are neurons [69,70]. In the brain of one genus of squid, *Loligo*, at least 30 distinct nucleus-like lobes have been identified [71]. The optic lobe, the largest of the fused central ganglia, contains as many as 65 million neurons. In addition to processing visual input, the optic lobe plays a critical role in higher motor control and the establishment of memory [69]. A number of other lobes may be functionally equivalent to vertebrate forebrain structures [69], though their organization bears little resemblance to the

Box 2. Animal selves

Animal selfhood can be conceived at many levels. A primitive or 'core' self may depend on self-modeling processes that use sensorimotor predictions to guide behavior leading to the emergence of a 'point-of-view' [13]. Core selfhood may arise from the complexity of the sensorimotor coordination and proprioception needed to support adaptive behavior [77]. It may also depend on 'feelings' or emotional states mediated by interoceptive representations of bodily states [92,93] related to homeostatic control of low-level drives for (at least) air, water, food, and pain avoidance [94,95].

Advanced selfhood involves the emergence of the subjective 'I' of higher order consciousness, and, in its most elaborated form, the ability to understand the world from the perspective of another. Basic requisites for a subjective 'I' are suggested by 'mirror self-recognition' (MSR) experiments, in which, for example, animals spontaneously use a mirror to examine an otherwise inaccessible body region [96]. MSR has been demonstrated in primates, including chimpanzees, orangutans, and (less conclusively) gorillas [96,97]; evidence suggestive of MSR has also been elicited from Asian elephants [98], dolphins [99], and magpies [100]. Even in chimpanzees, however, the prevalence of MSR behavior is only about 75% [101].

Even though animals lacking higher order consciousness cannot construct a subjective 'I,' taking humans as a benchmark suggests new experimental approaches. For example, combining virtual reality with tactile feedback can lead to a displacement of the first person perspective to a place outside the physical body, similar to spontaneous 'out-of-body' experiences [102]. This displacement, possibly due to a disruption of self-modeling, can be validated both by accurate report (e.g. 'I see myself from behind'), and by indirect behavioral evidence (e.g. subjects move in directions reflecting their perceived location). Similar results obtained from animals would suggest that they, too, possess a body-centered locus of experience.

Recent neuroimaging results have exposed new self-related neural processes in humans that may prompt similar searches in animals. Activity in the human 'default network' is correlated with stimulus-independent thought and self-related conscious content [103], and is anti-correlated with sensitivity to external somatosensory stimuli [104]. The right insular cortex appears important for self-consciousness [105], with anterior subregions possibly supporting explicit representations of feeling states underlying higher-order self-representations [92]. Hypoactivation of insula has been reported in humans with 'depersonalization disorder,' involving reduced subjective validity of the self [106]. Identification of animal analogs to these processes would require demonstration of very close homologies so far not achieved. Indeed, the thalamocortical pathway conveying interoceptive signals to the right anterior insula appears to be unique to primates [92], suggesting important species-specific differences as well as commonalities to be kept in mind when considering animal consciousness.

laminar sheets of mammalian cortex. In particular, the vertical, superior frontal, and inferior frontal lobes of octopus, squid, and cuttlefish are involved in memory consolidation [63,72,73]. In experiments in which the vertical lobe of *Octopus vulgaris* was lesioned, the ability to learn visual discriminations was severely impaired, but long-term memory consolidation remained intact [63]. Removal of the median inferior frontal lobe caused memory deficits that compromised learning [54,74]. Taken together, these studies suggest that some regions of the octopus frontal and vertical lobes are functionally comparable to regions of mammalian cortex (see Young [75] for a review).

Radical differences between cephalopod nervous systems and those of vertebrates are exemplified by the parallel, distributed architecture of the octopus locomotor system. The number of neurons in the tentacles of the octopus collectively exceeds the total number in the central fused ganglia of the brain itself [70]. A detached octopus arm will ail in a realistic manner when stimulated with short electrical pulses [76], suggesting pseudoautonomous control of some locomotor behavior patterns and hinting at a sophistication of sensorimotor coordination rivaling that of many vertebrates. This elaborate bodily representation in the service of sensorimotor coordination for adaptive behavior (e.g. locomotion, camouflage, mimicry) might support a 'core selfhood' [77] (see Box 2), a tantalizing concept as applied to cephalopods.

With regard to neuropharmacological and physiological properties, the cephalopod nervous system contains many of the major neurotransmitters that are found in mammalian brains [69,78]. In particular, the presence of dopamine (DA), noradrenaline (NA), and serotonin (5-HT) receptor subtypes that resemble those found in vertebrates may reflect the presence of circuitry similar to vertebrate excitatory and inhibitory pathways. As is the case with functional avian neuroanatomy, application of immunohistochemical and genetic techniques may help to determine the cephalopod functional analogs to neural regions in mammals that show correlated activity during conscious behavior. An encouraging indication is the recent identification of a cephalopod ortholog to the *Foxp2* gene, which in birds and humans has been implicated in motor function related to song and language production, respectively. Notably, *Foxp2* expression has been observed in the adult octopus chromatophore lobes [79].

What can be said of neurodynamics in cephalopod brains? Examination of octopus vertical lobe slices has identified long-term potentiation (LTP) of glutamatergic synaptic field potentials similar to those found in vertebrates [80]. More directly related to possible conscious states, electrophysiological studies have identified EEG patterns, including event related potentials, which resemble those of awake vertebrates, and at the same time are distinct from those recorded in other invertebrates [81,82]. Identifying cephalopod EEG patterns that reflect low amplitude fast irregular activity similar to that observed during human conscious states will require determination of suitable recording sites. Optic, vertical, and superior lobes of the octopus brain – all of which are critical to learning and memory – are relevant candidates.

The similarities discussed above by no means confirm the existence of conscious states in cephalopod molluscs, but neither do they exclude them. An intermediate effort to clarify the situation might be the pursuit of psychophysics in cephalopods, an approach not yet represented in the literature (see Box 1).

Concluding remarks

Approaches to animal consciousness require both clear theoretical frameworks and relevant experimental evidence. We have suggested that a useful approach is to synthesize neuroanatomical, neurophysiological, and behavioral evidence, using humans as a benchmark. We recognize that a distinction between primary and higher-order consciousness implies that mechanisms underlying putative primary conscious states might be distinct from, though possibly overlapping with, mechanisms allowing its accurate report, such that absence of evidence need not be evidence of absence in regard to animal consciousness.

Within this framework we find that birds exhibit rich cognitive and behavioral capabilities consistent with conscious states, including working memory, social learning, planning, and possibly even insight during problem solving. These capabilities are complemented by substantial anatomical homologies and functional similarities with mammals in the thalamocortical systems that are associated with consciousness. The case for cephalopod molluscs is currently much less clear. However, abundant evidence of sophisticated learning and memory faculties and rich behaviors, as well as early indications from studies of cephalopod neurophysiology, suggest at least the possibility of conscious states.

Given the profound gaps that remain in the neuroanatomical characterization of both subphyla, many basic questions remain concerning the existence, form, and prevalence of non-mammalian consciousness (see Box 3). Future progress in addressing these questions will require elaboration of behavioral paradigms designed to assess complex discriminatory behavior associated with consciousness. Vocal learning in birds provides a particularly promising avenue for achieving this objective (Box 1). In all cases, theoretical developments are necessary to facilitate the transition from correlation to causal explanation [14]. Such a transition will allow attribution of animal consciousness to be based causally on neural properties rather than on indirect behavioral report. Finally, we note that work on animal consciousness may help in assessing

Box 3. Outstanding questions

- How can we distinguish the neural mechanisms underlying accurate report from those underlying primary consciousness?
- Can degrees of consciousness be established among different species?
- What invertebrate neural structures and mechanisms may be analogous in function to mammalian thalamocortical systems?
- What function does consciousness serve in adaptive behavior?
- When did consciousness appear in the course of evolution?
- How might a synthetic approach facilitate the study of consciousness in other phyla (i.e. reptiles, amphibians, and bony and cartiligous fish)?
- What ethical implications would emerge if convincing evidence is obtained for widespread animal consciousness?

consciousness in humans incapable of report, including infants and patients in vegetative and minimally conscious states.

Acknowledgements
We are grateful to Drs. Graziano Fiorito, Joseph Gally, and Michael Beaton for their critical reading of the manuscript. We also thank Dr. Fiorito and Dr. Erich Jarvis for generously providing material used to design the cartoons of octopus and avian nervous systems, respectively, presented in Figure 2. Finally, we are indebted to Dr. Heidi Junger for her keen insights and assistance in the design of both figures. The work of D.B.E. is supported in part by DARPA (ONR Grant N00014-08-1-0728) and by the Neurosciences Research Foundation. A.K.S. is supported by EPSRC Leadership Fellowship EP/G007543/1.

References
1 Darwin, C. [2004 (1871)] The Descent of Man, and Selection in Relation to Sex. Penguin Classics
2 Griffin, D.R. (2000) Scientific approaches to animal consciousness. *Amer Zool* 4, 889–892
3 Edelman, G.M. (1989) *The Remembered Present.* Basic Books
4 Seth, A.K. and Baars, B.J. (2005) Neural Darwinism and consciousness. *Conscious Cogn.* 14, 140–168
5 Seth, A.K. *et al.* (2005) Criteria for consciousness in humans and other mammals. *Conscious Cogn.* 14, 119–139
6 Edelman, G.M. (2003) Naturalizing consciousness: a theoretical framework. *Proc. Natl. Acad. Sci. U. S. A.* 100, 5520–5524
7 Tononi, G. and Koch, C. (2008) The neural correlates of consciousness: an update. *Ann. N. Y. Acad. Sci.* 1124, 239–261
8 Baars, B.J. (1988) A cognitive theory of consciousness. *Cambridge University Press*
9 Block, N. (1995) On a confusion about a function of consciousness. *Behav. Brain Sci.* 18, 227–247
10 Knight, D.C. *et al.* (2006) The role of awareness in delay and trace fear conditioning in humans. *Cogn. Affect. Behav. Neurosci.* 6, 157–162
11 Tononi, G. and Edelman, G.M. (1998) Consciousness and complexity. *Science* 282, 1846–1851
12 Seth, A.K. (2009) Functions of consciousness. In *Encyclopedia of Consciousness* (Vol. 1) (Banks, W.P., ed.), In pp. 279–293, Academic Press
13 Metzinger, T. (2003) *Being No-One.* MIT Press
14 Seth, A.K. (2009) Explanatory correlates of consciousness: Theoretical and computational challenges. *Cogn. Computat.* 1, 50–63
15 Owens, A.M. *et al.* (2006) Detecting awareness in the vegetative state. *Science* 313, 1402
16 Nitz, D.A. *et al.* (2002) Electrophysiological correlates of rest and activity in Drosophila melanogaster. *Curr. Biol.* 12, 1934–1940
17 Raizen, D.M. *et al.* (2008) Lethargus is a Caenorhabditis elegans sleep-like state. *Nature* 451, 569–572
18 Siegel, J.M. (2008) Do all animals sleep? *Trends Neurosci.* 31, 208–213
19 Logothetis, N.K. (1998) Single units and conscious vision. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 353, 1801–1818
20 Srinivasan, R. *et al.* (1999) Increased synchronization of neuromagnetic responses during conscious perception. *J. Neurosci.* 19, 5435–5448
21 Gould-Beierle, K. (2000) A comparison of four corvid species in a working and reference memory task using a radial maze. *J. Comp. Psychol.* 114, 347–356
22 Pravosudov, V.V. and de Kort, S.R. (2006) Is the western scrub-jay (Aphelocoma californica) really an underdog among food-caching corvids when it comes to hippocampal volume and food caching propensity? *Brain Behav. Evol.* 67, 1–9
23 Emery, N.J. and Clayton, N.S. (2004) The mentality of crows: convergent evolution of intelligence in corvids and apes. *Science* 306, 1903–1907
24 Bugnyar, T. and Heinrich, B. (2006) Pilfering ravens, Corvus corax, adjust their behaviour to social context and identity of competitors. *Anim. Cogn.* 9, 369–376
25 Tumer, E.C. and Brainard, M.S. (2007) Performance variability enables adaptive plasticity of 'crystallized' adult birdsong. *Nature* 450, 1240–1244
26 Brainard, M.S. and Doupe, A.J. (2000) Auditory feedback in learning and maintenance of vocal behaviour. *Nat. Rev. Neurosci.* 1, 31–40
27 Pepperberg, I.M. (2002) In search of king Solomon's ring: cognitive and communicative studies of Grey parrots (Psittacus erithacus). *Brain Behav. Evol.* 59, 54–67
28 Pepperberg, I.M. and Lynn, S.K. (2000) Possible levels of animal consciousness with reference to grey parrots (Psittacus erithacus). *Integr. Comp. Biol.* 40, 893–901
29 Diekamp, B. *et al.* (2002) Nonspatial and subdivision-specific working memory deficits after selective lesions of the avian prefrontal cortex. *J. Neurosci.* 22, 9573–9580
30 Butler, A.B. (2008) Evolution of brains, cognition, and consciousness. *Brain Res. Bull* 75, 442–449
31 Lipp, H.P. *et al.* (2001) A large outdoor radial maze for comparative studies in birds and mammals. *Neurosci. Biobehav. Rev.* 25, 83–99
32 Akins, C.K. *et al.* (2002) Imitative learning in Japanese quail (Coturnix japonica) using the bidirectional control procedure. *Anim. Learn. Behav.* 30, 275–281
33 Campbell, F.M. *et al.* (1999) Stimulus learning and response learning by observation in the European starling, in a two-object/two-action test. *Anim. Behav.* 58, 151–158
34 Zentall, T.R. (2004) Action imitation in birds. *Learn. Behav.* 32, 15–23
35 Heyes, C. (2001) Causes and consequences of imitation. *Trends Cogn. Sci.* 5, 253–261
36 Bugnyar, T. and Heinrich, B. (2005) Ravens, Corvus corax, differentiate between knowledgeable and ignorant competitors. *Proc. Biol. Sci.* 272, 1641–1646
37 White, S.A. *et al.* (2006) Singing mice, songbirds, and more: models for FOXP2 function and dysfunction in human speech and language. *J. Neurosci.* 26, 10376–10379
38 Pepperberg, I.M. and Shive, H.R. (2001) Simultaneous development of vocal and physical object combinations by a Grey parrot (Psittacus erithacus): bottle caps, lids, and labels. *J. Comp. Psychol.* 115, 376–384
39 Farries, M.A. and Perkel, D.J. (2002) A telencephalic nucleus essential for song learning contains neurons with physiological characteristics of both striatum and globus pallidus. *J. Neurosci.* 22, 3776–3787
40 Durand, S.E. *et al.* (1997) Vocal control pathways through the anterior forebrain of a parrot (Melopsittacus undulatus). *J. Comp. Neurol.* 377, 179–206
41 Aboitiz, F. (1999) Comparative development of the mammalian isocortex and the reptilian dorsal ventricular ridge. Evolutionary considerations. *Cereb. Cortex* 9, 783–791
42 Farries, M.A. (2004) The avian song system in comparative perspective. *Ann. N. Y. Acad. Sci.* 1016, 61–76
43 Reiner, A. *et al.* (2005) Organization and evolution of the avian forebrain. *Anat. Rec. A Discov. Mol. Cell. Evol. Biol.* 287, 1080–1102
44 Karten, H.J. (1997) Evolutionary developmental biology meets the brain: the origins of mammalian cortex. *Proc. Natl. Acad. Sci. U. S. A.* 94, 2800–2804
45 Laverghetta, A.V. *et al.* (2006) Cellular localization of AMPA type glutamate receptor subunits in the basal ganglia of pigeons (Columba livia). *Brain Behav. Evol.* 67, 10–38
46 Reiner, A. *et al.* (2004) An immunohistochemical and pathway tracing study of the striatopallidal organization of area X in the male zebra finch. *J. Comp. Neurol.* 469, 239–261
47 Puelles, L. *et al.* (2000) Pallial and subpallial derivatives in the embryonic chick and mouse telencephalon, traced by the expression of the genes Dlx-2, Emx-1, Nkx-2.1, Pax-6, and Tbr-1. *J. Comp. Neurol.* 424, 409–438
48 Luo, M. and Perkel, D.J. (2002) Intrinsic and synaptic properties of neurons in an avian thalamic nucleus during song learning. *J. Neurophysiol.* 88, 1903–1914
49 McCormick, D.A. and Huguenard, J.R. (1992) A model of the electrophysiological properties of thalamocortical relay neurons. *J. Neurophysiol.* 68, 1384–1400
50 Rattenborg, N.C. *et al.* (2009) Avian sleep homeostasis: Convergent evolution of complex brains, cognition and sleep functions in mammals and birds. *Neurosci. Biobehav. Rev.* 33, 253–270
51 Butler, A.B. and Cotterill, R.M. (2006) Mammalian and avian neuroanatomy and the question of consciousness in birds. *Biol. Bull.* 211, 106–127

52 Mather, J.A. (2008) Cephalopod consciousness: behavioural evidence. *Conscious Cogn.* 17, 37–48
53 Borrelli, L. and Fiorito, G. (2008) Behavioral analysis of learning and memory in cephalopods. In *Learning Theory and Behavior* (Menzel, R., ed.), pp. 605–628, Elsevier
54 Wells, M.J. and Young, J.Z. (1972) The median inferior frontal lobe and touch learning in the octopus. *J. Exp. Biol.* 56, 381–402
55 Sutherland, N.S. (1969) Shape discrimination in rat, octopus, and goldfish: a comparative study. *J. Comp. Physiol. Psychol.* 67, 160–176
56 Fiorito, G. *et al.* (1998) The role of stimulus preexposure in problem solving by Octopus vulgaris. *Anim. Cogn.* 1, 107–112
57 Fiorito, G. *et al.* (1990) Problem solving ability of Octopus vulgaris Lamarck (Mollusca, Cephalopoda). *Behav. Neural Biol* 53, 217–230
58 Fiorito, G. and Scotto, P. (1992) Observational learning in Octopus vulgaris. *Science* 256, 545–547
59 Biederman, G.B. and Davey, V.A. (1993) Social learning in invertebrates. *Science* 259, 1627–1628
60 Suboski, M.D. *et al.* (1993) Social learning in invertebrates. *Science* 259, 1628–1629
61 Ludwig, K.R. *et al.* (1993) Response: Social learning in invertebrates. *Science* 259, 1626–1627
62 Fiorito, G. (1993) Response: Social learning in invertebrates. *Science* 259, 1629
63 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
64 Moriyama, T. and Gunji, Y.P. (1997) Autonomous learning in maze solution by octopus. *Ethology* 103, 499–513
65 Gil, M. *et al.* (2007) Learning reward expectations in honeybees. *Learn. Mem.* 14, 491–496
66 Menzel, R. (2001) Searching for the memory trace in a mini-brain, the honeybee. *Learn. Mem.* 8, 53–62
67 Jackson, R.R. *et al.* (2002) Interpopulation variation in the risk-related decisions of Portia labiata, an araneophagic jumping spider (Arianae, Salticidae), during predatory sequences with spitting spiders. *Anim. Cogn.* 5, 215–223
68 Jackson, R.R. and Wilcox, R.S. (1998) Spider-eating spiders. *Am. Sci.* 86, 350–357
69 Hanlon, R.T. and Messenger, J.B. (1996) *Cephalopod Behavior.* Cambridge University Press
70 Young, J.Z. (1963) The number and sizes of nerve cells in octopus. *Proc. Zool. Soc. Lond.* 140, 229–254
71 Maddock, L. and Young, J.Z. (1987) Quantitative differences among the brains of cephalopods. *J. Zool. Soc. Lond.* 212, 739–767
72 Young, J.Z. (1961) Learning and discrimination in the octopus. *Biol. Rev. Camb. Philos. Soc.* 36, 32–96
73 Williamson, R. and Chrachri, A. (2004) Cephalopod neural networks. *Neurosignals* 13, 87–98
74 Wells, M.J. and Young, J.Z. (1969) The effect of splitting part of the brain or removal of the median inferior frontal lobe on touch learning in octopus. *J. Exp. Biol.* 50, 515–526
75 Young, J.Z. (1995) Multiple matrices in the memory system of Octopus. In *Cephalopod Neurobiology* (Abbot, J.N., ed.), pp. 431–443, Oxford University Press
76 Sumbre, G. *et al.* (2001) Control of octopus arm extension by a peripheral motor program. *Science* 293, 1845–1848
77 Merker, B. (2005) The liabilities of mobility: a selection pressure for the transition to consciousness in animal evolution. *Conscious Cogn.* 14, 89–114
78 Messenger, J.B. (1996) Neurotransmitters of cephalopods. *Invertebrate Neurosci.* 2, 95–114
79 Sirakov, M. (2007) Identification and characterization of FoxP in the central nervous system of Octopus vulgaris (Mollusca, Cephalopoda) [Identifcazione e caratterizzazione di FoxP nel sistema nervoso centrale di Octopus vulgaris (Mollusca, Cephalopoda)], PhD Thesis, Università della Calabria
80 Shomrat, T. *et al.* (2008) The octopus vertical lobe modulates short-term learning rate and uses LTP to acquire long-term memory. *Curr. Biol.* 18, 337–342
81 Bullock, T.H. and Basar, E. (1988) Comparison of ongoing compound field potentials in the brains of invertebrates and vertebrates. *Brain Res.* 472, 57–75
82 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
83 Bachmann, T. *et al.* (2007) *Experimental Phenomena of Consciousness.* Oxford University Press
84 Cowey, A. and Stoerig, P. (1995) Blindsight in monkeys. *Nature* 373, 247–249
85 Schmidt, M.F. (2008) Using both sides of your brain: the case for rapid interhemispheric switching. *PLoS Biol.* 6, e269
86 Hanlon, R.T. and Messenger, J.B. (2002) *Cephalopod Behavior.* Cambridge University Press
87 Hanlon, R.T. *et al.* (2009) Cephalopod dynamic camouflage: bridging the continuum between background matching and disruptive coloration. *Philos. Trans. R. Soc. Lond. B Biol. Sc. i* 364, 429–437
88 Borrelli, L. *et al.* (2006) *A Catalogue of Body Patterning in Cephalopoda.* Firenze University Press
89 Messenger, J.B. (2001) Cephalopod chromatophores: neurobiology and natural history. *Biol. Rev. Camb. Philos. Soc.* 76, 473–528
90 Shashar, N. and Cronin, T.W. (1996) Polarization contrast vision in Octopus. *J. Exp. Biol.* 199, 999–1004
91 Shashar, N. *et al.* (1996) Polarization vision in cuttlefish in a concealed communication channel? *J. Exp. Biol.* 199, 2077–2084
92 Craig, A.D.B. (2009) How do you feel – now? The anterior insula and human awareness. *Nat. Rev. Neurosci.* 10, 59–70
93 Damasio, A. (2000) *The Feeling of What Happens: Body, Emotion and the Making of Consciousness.* Vintage
94 Panksepp, J. (2005) Affective consciousness: Core emotional feelings in animals and humans. *Conscious Cogn.* 14, 30–80
95 Denton, D.A. *et al.* (2008) The role of primordial emotions in the evolutionary origin of consciousness. *Conscious Cogn.* 18, 500–514
96 Suddendorf, T. and Collier-Baker, E. (2009) The evolution of primate visual self-recognition: evidence of absence in lesser apes. *Proc. Biol. Sci.* 276, 1671–1677
97 Posada, S. and Colell, M. (2007) Another gorilla (Gorilla gorilla gorilla) recognizes himself in a mirror. *Am. J. Primatol.* 69, 576–583
98 Plotnik, J.M. *et al.* (2006) Self-recognition in an Asian elephant. *Proc. Natl. Acad. Sci. U. S. A.* 103, 17053–17057
99 Reiss, D. and Marino, L. (2001) Mirror self-recognition in the bottlenose dolphin: a case of cognitive convergence. *Proc. Natl. Acad. Sci. U. S. A.* 98, 5937–5942
100 Prior, H. *et al.* (2008) Mirror-induced behavior in the magpie (Pica pica): evidence of self-recognition. *PLoS Biol.* 6, e202
101 Povinelli, D.J. *et al.* (1993) Self-recognition in chimpanzees (Pan troglodytes): distribution, ontogeny, and patterns of emergence. *J. Comp. Psychol.* 107, 347–372
102 Lenggenhager, B. *et al.* (2007) Video ergo sum: manipulating bodily self-consciousness. *Science* 317, 1096–1099
103 Mason, M.F. *et al.* (2007) Wandering minds: the default network and stimulus-independent thought. *Science* 315, 393–395
104 Boly, M. *et al.* (2007) Baseline brain activity fluctuations predict somatosensory perception in humans. *Proc. Natl. Acad. Sci. U. S. A.* 104, 12187–12192
105 Critchley, H.D. *et al.* (2004) Neural systems supporting interoceptive awareness. *Nat. Neurosci.* 7, 189–195
106 Phillips, M.L. *et al.* (2001) Depersonalization disorder: thinking without feeling. *Psychiatry Res.* 108, 145–160