REVISED MANUSCRIPT

Non-human consciousness and the specificity problem: a modest theoretical proposal

ABSTRACT. The scientific study of consciousness has yielded a number of constructive frameworks concerning the cognitive basis of consciousness in humans. However, these theories are challenging to apply to non-human systems insofar as few if any non-humans are likely to implement the precise cognitive architecture of human beings, an issue I term the Specificity Problem. In this paper, I explore possible solutions to this challenge. After providing some background on the theories of consciousness debate and spelling out the Specificity Problem, I survey four approaches to the challenge that I term Conservatism, Liberalism, Incrementalism, and Rejectionism. I then consider a possible solution to the Specificity Problem, namely the Theory-Light approach proposed by Jonathan Birch. This holds that we avoid unnecessary theoretical commitments at the outset of inquiry and instead investigate nonhuman consciousness using behavioural markers. After noting limitations of this method, I argue for what I term a Modest Theoretical Approach. This holds that we can use the insights of the Theory-Light approach in combination with theoretical methods to resolve the Specificity Problem. Via this two-pronged strategy, I suggest, we can gain insights into challenging cases that would be intractable using a Theory-Light method alone.

9478 words

1. Introduction

The science of consciousness has made considerable progress on multiple fronts, including the development of a variety of theoretical approaches and the refinement of techniques for measuring different forms of awareness. A key outstanding challenge, however, concerns how we can assess consciousness in non-human systems. At least in principle, we might hope that the same theories of consciousness that aim to predict conscious awareness in humans could also tell us whether beings like rats, fish, and insects, as well as artificial systems, are conscious.

In this paper, I explore a challenge faced by many current theories of consciousness concerning such attributions of consciousness to non-humans that I term the *Specificity Problem*. In short, this is the problem of how we can idenitfy the appropriate level of detail that should be adopted by theories of consciousness in applying them to non-humans. If we spell out our theories in a way that makes detailed reference to specific aspects of human cognitive architecture, we run the risk of false negatives, inappropriately denying consciousness to non-human systems in virtue of minor differences in processing. However, if we spell out our theories in very abstract terms, we face the opposite worry, namely misclassifying non-conscious systems as conscious entities.

The paper proceeds as follows. I begin in Section 2 by providing a brief background of the theories of consciousness debate, before going on in Section 3 to summarise the Specificity Problem, explaining how it presents an obstacle for the application of views such as Global Workspace Theory to non-human systems. In Section 4, I discuss four initial responses to the Specificity Problem that I term *Conservatism, Minimalism, Incrementalism,* and *Rejectionism*. In Section 5, I explore a distinction (due to Birch, 2019) between *Theory-Heavy* and *Theory-Light* approaches to consciousness, and suggest that the Specificity Problem only poses a significant problem if we are committed to the former methodology. However, I also note that a Theory-Light approach is unlikely to provide all of the answers we need from a science of non-human consciousness. Finally, in Section 6, I propose an alternative strategy for addressing the Specificity Problem that I term the *Modest Theoretical Approach*. I claim that by combining methods of both Theory-Heavy and Theory-Light approaches, we can resolve the Specificity Problem and make determinations of consciousness in difficult cases, as well as gaining some insights into which theories of consciousness have the greatest promise.

2. The theories of consciousness debate

A central goal in the study of consciousness has been to develop an account to explain why some forms of psychological and neural processing are conscious and others are not. Frameworks that attempt to offer such explanations are typically known as *theories of consciousness*, and the business of assessing and critiquing such theories as the *theories of consciousness debate*.[1] While the theories of consciousness debate draws on longstanding philosophical and scientific ideas, as a specialised area of cognitive science involving active collaboration between philosophers and scientists, its modern history arguably began in the 1980s, with the emergence of frameworks such as Rosenthal's Higher-Order Thought theory (Rosenthal 1986) and Baars' Global Workspace account (1988). Since then, a host of other theories have developed, with notable examples including Tononi's Integrated Information Theory (Tononi, 2008), biological theories of the kind defended by Ned Block and Victor Lamme (Ned Block, 2009; Lamme, 2010), and Dehaene and Naccache's Global Neuronal Workspace (Dehaene & Naccache, 2001).

A key method used by researchers in the field involves examining the conditions under which different stimuli are consciously or unconsciously perceived by human subjects. Paradigms of this kind include binocular rivalry (Blake & Logothetis, 2002), masking (Kouider & Dehaene, 2007), and dichotic listening (Brancucci & Tommasi, 2011), all of which rely on the fact that by varying conditions of input or task, we can control whether subjects can detect and report on the presence of a stimulus. By identifying the cognitive and neural dynamics associated with detectability and reportability – phenomena frequently though controversially associated with consciousness – we can attempt to construct accounts of what distinguishes conscious from unconscious forms of processing, and from there give a general theory of

[1] A key distinction in the debate is that between phenomenal consciousness (understood as subjective experience) and access consciousness (the ability to deploy information in voluntary action, including report) (Block, 1995) . For ease of exposition and to avoid prejudging theoretical questions, I use the term "consciousness" to refer to subjective experience, while leaving open the possibility that phenomenal consciousness and access consciousness are coreferring terms.

consciousness.[2] Other important methods used in the science of consciousness include experimental work on patients with brain damage that causes selective impairment of consciousness (Weiskrantz, 1986) and brain-imaging of patients in conscious and unconscious conditions such as general anaesthesia (Alkire & Miller, 2005).

Considerable controversy remains concerning which if any of the leading theories of consciousness is to be preferred. However, the field as a whole has proven highly productive, giving rise to a wealth of data about different forms of processing in the brain, with implications for the detection of consciousness in vegetative state patients (Casali et al., 2013), and has led to the development of key theoretical concepts such as Block's distinction between phenomenal and access consciousness (Block 1995).

Despite these advances in understanding the basis of consciousness in humans, however, comparatively little progress has been made within the debate concerning non-human consciousness.[3] One important reason why this has been the case is that that researchers in this area overwhelmingly rely on *verbal report* to provide clear evidence of the fact that a given stimulus has been consciously processed. While some efforts have been made to move beyond verbal report (for example, Tsuchiya, Wilke, Frässle, & Lamme, 2015), these remain controversial. As a result, it is difficult to apply many of the central experimental paradigms to debates about consciousness in non-humans.[4] Likewise, some approaches (such as that of Crick & Koch, 1990, and Lamme, 2010) are primarily concerned with identifying neural and biological mechanisms of consciousness in humans, and are thus not intended to offer an account of consciousness that can be generalised to non-human systems.

Nonetheless, many of the leading theories of consciousness are more ambitious, seeking to provide an analysis of consciousness *tout court* in terms of some form of psychological or cognitive mechanism such as metacognition or global broadcast. Consequently, such theories might expect to be able to answer the question of which non-

[2] As suggested in the text, there are major ongoing debates concerning the relationship between consciousness and reportability. See Block (2007) for a useful summary.

[3] More extensive discussion of non-human consciousness has of course occurred in other branches of cognitive science, notably cognitive ethology and the philosophy of animal minds (see, e.g., Bekoff, 2010; Godfrey-Smith, 2017; Griffin, 2001). See Section 5 for further discussion.

[4] Some notable exceptions include binocular rivalry experiments in monkeys (Stoerig et al., 2002) and delayed trace eyeblink conditioning in rabbits (Thompson et al., 1996).

human systems are conscious by determining which of them implement the relevant processes. Putting this into practice for any given non-human system will of course present many challenges and will require us to conduct painstaking investigation to determine the details of the system's cognitive architecture, but our over-arching methodology, at least, will be fairly straightforward.

The idea that we could use our theories of human consciousness to settle questions about non-human consciousness in such a direct way is controversial (see Section 5, below, for discussion of this 'Theory-Heavy' approach). However, I take it that many of the leading theories – including Global Workspace Theory (hence, GWT;(Dehaene & Naccache, 2001), working memory theories (Baddeley, 1992), higher-order thought theories (Rosenthal, 2005), Prinz's Attended Intermediate Representations (AIR) view (Prinz, 2012), and the Attention Schema theory (Graziano, 2013) – lend themselves to this strategy, and I will refer to them collectively as *cognitive* theories of consciousness. Such theorists, I will now argue, face a challenge I term the Specificity Problem.

3. The Specificity Problem

In short, the Specificity Problem is the challenge of how to spell out the cognitive mechanisms identified as constitutive of consciousness according to our preferred theory in such a way as to make them applicable beyond the human case. Processes such as global activation, working memory, attention, and metacognition are all likely to manifest in non-human systems in ways that differ more or less dramatically from the human case. The difficulty is how we can identify which of these differences matter for questions about consciousness and which can be set aside.

I should note at the outset that the Specificity Problem as sketched above is not novel. Worries about finding the right balance between 'liberalism' and 'chauvinism' in our attributions of mentality and consciousness to non-human systems, for example, have long dogged functionalist approaches to the mind (Block, 1978), and several philosophers have engaged with similar challenges for specific theories of consciousness. Hence Prinz (2018) in considering the applicability of his AIR theory to reptiles and amphibians notes that "their information-processing systems are similar to ours in certain respects, but also different in ways that reduce confidence when attributing consciousness." Likewise, in attempting to apply GWT to garden snails, Schwitzgebel (2019) notes that "information travels broadly through snail brains, enabling coordinated action" and asks whether this is "global workspace enough" to count as evidence for consciousness. Most notably, perhaps, in recent work Carruthers (2018) raises a version of the Specificity Problem faced by his preferred account of human consciousness, namely GWT, concluding that it renders questions of animal consciousness meaningless (see 4.4, below).[5]

What I take to be both novel and important in the present discussion, however, is the significance of the Specificity Problem as a *general* problem for cognitive theories of consciousness in their application to any non-human system. As an initial illustration, consider Dehaene and Naccache's Global Workspace Theory (hence, GWT). At times, this theory is spelled out in highly abstract terms, for example in the claim that "consciousness is just brain-wide information sharing" (Dehaene, 2014: 165). However, elsewhere Dehaene seems to have a much narrower view in mind, asserting, for example, that "the capacity to report is a key feature of a conscious state." This prompts the question of which capacities are strictly *critical* for global broadcast in the sense relevant for consciousness, and how much a system's architecture can diverge from the human case while still being conscious. Consider, for example, that the human global workspace makes information available for a wide variety of cognitive functions such as report, self-awareness, and long-term planning. Which if any of these are necessary for a system to possess a global workspace in the strict sense is by no means obvious. As Carruthers (2018) puts it, "[a]re some aspects of global broadcasting more important, or more relevant to the question of consciousness, than others?"

Versions of this problem arguably arise for all cognitive theories of consciousness insofar as they appeal to broadly specified cognitive functions that can be realised in a host of different ways in different biological or artificial systems. Consider, for example, the higher-order approach to consciousness which claims that consciousness essentially involves awareness of our own mental states. This might take the form of some form of internal quasi-perceptual awareness of our own mental states (Lycan, 1996) or higher-

[5] In addition to these instances of the Specificity Problem in consciousness research, there are, of course, many related instances in comparative philosophy and cognitive science where methodological issues are raised by differences between humans and non-human animals. see, e.g., Buckner (2015) for the case of cognition, Carruthers (2013) for working memory, and Clayton et al. (2007) for episodic and 'episodic-like' memory.

order thoughts (Rosenthal, 2005). In order to apply such theories to non-human systems, however, we must identify *what counts* as a perception or a thought in the required sense, and here arise complex conceptual questions. Metacognition – broadly understood as a capacity to internally represent one's own mental states – comes in a host of different forms and is found to varying degrees in different non-human systems. At its simplest, this might involve merely differentiating between internally and externally generated changes in perceptual inputs (Merker, 2005), but the concept also includes more sophisticated capacities such as recognising that others can hold beliefs quite different from one's own (Kaminski et al., 2008) and being able to represent one's degree of confidence in different beliefs (Terrace & Son, 2009). In all cases, the cognitive resources available to non-humans to conceptualise and structure their metacognitive representations are likely to be very different from those possessed by human beings. For example, via our linguistic capabilities and our rich repertoire of concepts, a human can represent herself as having thoughts with complex logical structure. Whether such complex forms of self-representation are required to possess higher-order thoughts in the sense relevant for consciousness is thus not trivially answered.

I do not wish to overstate the difficulties in overcoming the Specificity Problem. It is reasonable to hope that there is *some* appropriate level of specificity in our theories of consciousness that allows us to make accurate attributions of consciousness to non-human systems. The challenge is how we can identify it. A key point to note in this regard is that the experimental methods used to develop our theories of consciousness in the first place may be ill-suited to the task. We may be able to determine via techniques such as masking and binocular rivalry that activation of the global workspace suffices for consciousness in humans, for example, but the question of what we should *count* as a global workspace in non-human cases will require different forms of theorising and investigation.

4. Four solutions to specificity

In Sections 5 and 6, I will outline what I take to be a promising methodology for calibrating the specificity of our theories of consciousness so as to allow them to be accurately applied to non-humans. First, however, it will be helpful to map some of the broader theoretical terrain. With this in mind, I will now briefly survey four possible approaches, identifying what I take to be some initial difficulties with each. I would stress, however, that the brief remarks that follow are not intended as knockdown objections. Moreover, I do not take the approaches considered below to be exhaustive of solutions to the Specificity Problem, nor even necessarily to constitute the best or most sophisticated answers we could adopt. They have been selected rather because they attempt relatively clear and straightforward resolutions of the issue, and thus help to provide an outline for the broader space of possible answers.

4.1 – Conservatism

One straightforward strategy for applying our theories of consciousness to non-human systems would be to err on the side of conservatism, and assume – absent clear evidence to the contrary – that the full range of capacities associated with our candidate mechanism for consciousness in humans is essential. A Conservative reading of GWT, for example, might begin with the observation that globally broadcast information in humans can be employed in reflective thought, deliberation, self-awareness, and verbal report, and identify all of these capacities as necessary for consciousness. By contrast, a Conservative Higher-Order Thought theorist might require that any conscious creature possess a capacity for complex conceptually structured self-reflective thought. This would provide a clear answer to the Specificity Problem, insofar as consciousness would be confined just to those systems possessing very close analogues of the cognitive mechanisms identified as critical for consciousness in humans.

It is likely to follow from a Conservative view that few if any non-human animals are conscious, on the simple basis that humans have a number of seemingly cognitive endowments seemingly unique in the natural world, such as our sophisticated linguistic and conceptual capabilities. This should not lead us to dismiss it out of hand, of course. The idea that consciousness and higher cognitive functions are closely linked and that non-human animals may consequently lack consciousness altogether is one that has had many defenders both contemporary and historical (see, e.g., Carruthers, 1999; Davidson, 1975).

Nonetheless, few modern readers will be likely wish to endorse a view that circumscribes consciousness to humans alone. While we might debate how much evidential weight if any should be given to these intuitive considerations, a perhaps more pressing worry for the simple form of Conservatism sketched above is that such a position would be unduly anthropocentric, not insofar as it denies consciousness to non-human systems per se, but rather that it seems unmotivated to rule out consciousness in systems with cognitive architectures different from our own *just* on the basis of these differences. It is of course entirely possible that *some* differences in cognitive architecture will be critical for whether or not a system is conscious, but we should not assume at the outset that *all* such differences will be relevant.

To illustrate the point, I would suggest that it is not outlandish to suppose that a being might exist whose intelligence and behavioural capacities were broadly equivalent to that of humans, yet which differed from us considerably in respect of its cognitive architecture, perhaps possessing a system of internal representations more map-like than conceptual (Boyle, 2019; Camp, 2007) or having short-term memory systems with significantly different properties from our own. While it could be the case that these differences render the system non-conscious, to outright dismiss the possibility of consciousness in such a system just by virtue of there obtaining *any* major architectural differences in its relevant cognitive mechanisms seems somewhat unmotivated, not to mention mysterious. Consequently, while I recognise that more might be said to make a Conservative view appealing, I would suggest that at least at first glance it seems an unpromising answer to the Specificity Problem.

4.2 – Liberalism

Another straightforward approach might identify consciousness with the cognitive capacities and mechanisms proposed by the theory in question at the greatest possible degree of generality. For GWT, this might mean any form of system-wide information sharing, while a very liberal metacognitive approach to consciousness might identify consciousness as present in any system able to use the properties of its own mental states as inputs to other cognitive processes (cf. Shea, 2012b).

While this Liberal approach avoids the vices of Conservatism, it is vulnerable to the opposite charge, namely that it is excessively *generous* in its attributions of consciousness. Phenomena like global information-sharing and metacognition in their most liberal interpretations are ubiquitous both in nature and in existing artificial systems, and likely to be able to be implemented independently of the broader cognitive capacities of a system such as its sensorimotor capacities, or general intelligence. To give a toy example, consider Network Time Protocol (NTP), a tool for synchronisation of internal clocks across devices within a network. On the face of it, at least, this might qualify as a form of global information sharing (at least according to some radically liberal interpretation of GWT), yet it is hard to take seriously the possibility that such a relatively simple process would give rise to consciousness.

This intuition is admittedly not universally shared; there are those who will welcome a view of consciousness as ubiquitous in natural and artificial systems. Prominent defenders of Integrated Information Theory, for example, regard consciousness as "likely widespread among animals, and… found in small amounts even in certain simple systems" (Tononi & Koch, 2015). While such views cannot be ruled out from the armchair, it is worth recognising that they do considerable violence to our pretheoretical views about the distribution of consciousness. Schwitzgebel puts the point well, noting that "[w]e know, we think, prior to our theory-building, that the range of conscious entities does not include protons or simple logic gates."

The significance of such pretheoretical platitudes is itself contested, but on at least one influential view (Lewis, 1972), it should be a guiding constraint in our search for a theory that it systematise as far as possible our pretheoretical judgements about which states and systems are conscious. Even if we do not wish to adopt this specific methodology, we will likely wish to give *some* weight to pretheoretical judgements about consciousness, not least in recognition of the role that such judgements play in identifying the initial phenomenon we wish to explore. [6] Most of us would feel comfortable dismissing out of hand, for example, a theory of consciousness that claimed that children were not conscious until their teenage years, or that nematode worms were conscious and dogs were not. Without taking our pretheoretical priors as sacrosanct, then, we might have good reason to give them some evidential role in assessing scientific theories of consciousness. Consequently, to the extent that a Liberal approach is grossly at odds with these priors – for example, in asserting that a very simple robot may be conscious – it will, ceteris paribus, be less attractive than one that provides a plausible account of consciousness without doing violence to our pretheoretical attitudes.

[6] See Smith (1994; Ch.2) for a discussion of this methodology as applied to meta-ethics. See also Boyd (1999) on the role of 'programmatic' and 'explanatory' definitions for natural kinds.

4.3 – Incrementalism

We might attempt to steer a course between Conservatism and Liberalism by adopting a third approach I will label Incrementalism. In short, I will use this term to refer to views that claim consciousness comes in degrees. The idea here is not merely that different systems will have different *kinds* of consciousness that vary in intensity, complexity, or richness (something that all theorists would likely agree upon), but rather that the basic property of consciousness itself is not an all or nothing matter. Rather than asking, for example, whether there *is* or *is not* something it's like to be a crab or a computer, we might ask *to what extent* there is something it's like to be such a system.

Incrementalism as sketched above is a position that one might be sympathetic towards or reject for reasons quite orthogonal to the present discussion. However, we might hope that having adopted such a position, the Specificity Problem can be handily dealt with. Hence we could combine an Incrementalist view with a commitment to Global Workspace Theory, and suggest that 'full consciousness' required the complete cognitive architecture of the human Global Workspace, while allowing that other systems might have some *degree* of consciousness to the extent that the system approximated that architecture. Alternatively, one might identify full consciousness with the rich metacognitive capacities of human beings, while allowing that creatures with more rudimentary forms of metacognition might still have consciousness to some lesser degree.

Incrementalist approaches have obvious attractions insofar as they allow us to cast a broad net as regards the range of systems that might qualify as conscious, while allowing us to maintain that in some important sense, a creature such as a lobster or simple robot is vastly *less* conscious than we are. They also arguably provide a good fit for an evolutionary approach to consciousness, insofar as they suggest how consciousness might have emerged gradually in forms very different from that we associate with human experience.

A fundamental obstacle for many in endorsing such a view, however, will be the very notion of degrees of consciousness. While it is certainly true that we undergo conscious experiences that differ in respect of clarity, vividness, and richness, as for example when waking up or falling asleep, or under the influence of drugs or alcohol, in all of these instances it is still the case that there is *something it is like* (Nagel, 1974) to have the relevant experience; that is, we are still unambiguously in a conscious state. As Carruthers puts it, “One can be semi-conscious, or only partly awake. But some of the states one is in when semi-conscious are fully (unequivocally) phenomenally conscious nonetheless” (see also Simon, 2017). Such familiar examples, then, do not help us make sense of the idea of what it would mean for a creature to have some (non-total) degree of consciousness. I do not wish to rest too much on this point, not least because the question of whether consciousness can come in degrees or exhibit vagueness is a deeply contested one among philosophers (Papineau, 2003; Rosenthal, 2019). Nonetheless, to the extent that the reader finds degrees of consciousness a mysterious notion, it will be detract from the appeal of an Incrementalist solution to the Specificity Problem.

Another worry for Incrementalism concerns how we can established principled criteria for assessing similarity across different non-human systems, some of which might possess radically different cognitive architectures from ours. If a system possesses propositionally structured thought, for example, yet lacks sensorimotor capacities, should we say that it is more or less conscious than one with the opposite endowments? The point is not that such comparisons are pointless or impossible, but rather that they can be made only on the basis of some further theoretical perspective on the relative importance of different cognitive capacities, and this perspective is not likely to be neatly provided by the specific theory of consciousness we wish to endorse. The basic framework of GWT, for example, does not by itself specify which of the myriad consumer systems of globally broadcast information are especially important for consciousness.

To address the question of which systems matter for consciousness, we might attempt to appeal to evidence from lesion studies in which some cognitive capacities are absent or dramatically impaired while consciousness seems to be broadly intact, such as the loss of episodic memory among patients with hippocampal lesions (see, e.g., Burgess, Maguire, & O’Keefe, 2002). However, such cases do not settle the question of the *degree* to which consciousness might be present. We might decide to prioritise verbal report, for example, and claim that patients still able to report on their concurrent experiences are fully conscious. But we could equally claim that, shorn of their ability to hold onto memories for more than a few seconds, patients with severe anterograde amnesia have a lower degree of consciousness than non-verbal patients who have fully intact episodic memory. Such questions are of a deep and largely theoretical kind, and seem unlikely to be answered just by reference to empirical data, and while I do not take this to count against Incrementalism as a broader perspective in the science of consciousness, it does suggest at least that it may not provide a straightforward solution to the Specificity Problem.

4.4 – Rejectionism

A final approach I wish to consider is what I will call *Rejectionism regarding* about the application of cognitive theories of consciousness to non-humans. By this, I mean the view that questions about non-human consciousness are ill-posed, not because non-human systems are straightforwardly unconscious but because the categories of conscious and unconscious cannot be intelligibly applied to them. This is the considered view of Carruthers (Carruthers, 2018a, 2018b), who argues – based on considerations similar to those raised in Section 3 above – that questions about non-human consciousness are misconceived and lack any informative or useful answer.

While it is not possible in the present context to give a detailed summary of Carruthers' argument, its main points can be condensed as follows. First, Carruthers claims that global broadcasting is a strong candidate mechanism for consciousness in humans. Second, he argues – contra Incrementalism - that human consciousness is "all or none", and that our mental states are either phenomenally conscious or they are not.[7] Third, he suggests that "global broadcasting admits of degrees across species", and (as I argued above) that there are no principled ways of determining which aspects of the Global Workspace in humans are critical and which incidental to our conscious experience.

In light of these considerations, he suggests that two positions regarding animal consciousness may follow, depending on whether we take global broadcast to be merely associated with consciousness or to constitute a fully reductive account. If we take the former "qualia realist" position, then "the question of animal consciousness becomes intractable… [f]or there is no way to know which components of the workspace are

[7] While Carruthers seems to defend this claim largely on conceptual grounds (for example, stating that "it is hard to conceive of a perceptual state that is partly phenomenally conscious, partly not"), he also notes that the neural dynamics of global broadcast also exhibit this 'all or none' characteristic: activations in neural populations either reach a threshold required for "global ignition" or they do not (Dehaene, 2014).

sufficient for the presence of qualia." By contrast, if we adopt the latter view, as Carruthers urges, then we must grant that the term "consciousness" picks out a specific property of human cognitive dynamics that few if any other animals will instantiate exactly insofar as they resemble and differ from us in myriad different ways across species. Consequently, Carruthers argues, "there is no fact of the matter about animal consciousness", and moreover, "the question of animal consciousness is no longer of any significant interest."[8]

Carruthers' arguments are sophisticated and developed across multiple papers and a forthcoming book (Carruthers, 2020). Nonetheless, its conclusions are undeniably radical – perhaps radical enough to lead many to assume that any inferences that establish them must be flawed. As Schwitzgebel (2019) puts it, "[w]e are more confident that there is something it is like to be a dog than we could ever be that a clever philosophical argument to the contrary was in fact sound."[9] For my part, while I reject Carruthers' view, I will not be able to provide an adequate response to it in what remains of the present paper. Consequently, and in light of this omission, I recognise that the question which frames the following discussion must be understood as conditional: on the assumption that Rejectionism is false and the matter of animal consciousness is indeed a substantive issue, what possible answer do we have for the Specificity Problem?

5. Theory-heavy and theory-light approaches to consciousness

My argument thus far has been that the Specificity Problem presents a major challenge for the application of cognitive theories of consciousness to non-human systems. In what follows, I wish to argue that while the problem may require us to revise some of our methods for approaching animal consciousness, it is not insoluble. In order to sketch a solution, it will first be necessary to discuss a key distinction between two broad

[8] A natural objection might be that questions about animal consciousness have irreducible *moral* interest, insofar as we have special obligations towards sentient creatures capable of feeling pain (see e.g., Regan, 2004; Singer, 1989). However, Carruthers has argued elsewhere (1999, 2005) that moral issues of animal welfare do not hinge on issues of consciousness but can be reframed in terms of frustrations of desire (see also Dawkins, 2012).

[9] Note also the famous (and controversial) Cambridge Declaration on Consciousness, signed in 2012 by many leading neuroscientists, which asserted that "the weight of evidence indicates that humans are not unique in possessing the neurological substrates that generate consciousness."

approaches to the problem of non-human consciousness, namely the ‘Theory-Heavy’ and ‘Theory-Light’ approaches (Birch, 2019).

In short, the Theory-Heavy approach is that which was tacitly assumed in the above discussion, namely that we should approach questions about animal consciousness by assessing whether they possess the relevant mechanisms identified by our best theory of human consciousness. According to this approach, the challenges we face in determining which non-human systems are conscious are, first, to identify the mechanisms that make individual mental states conscious in humans, and second, to identify which non-human systems implement the relevant mechanisms and thus possess at least some conscious states. This approach, as argued above, runs into a powerful challenge in the form of the Specificity Problem.

Faced with this challenge, we might instead look for a quite different strategy for assessing the presence of consciousness in non-human systems. One such method is what Birch terms the ‘Theory-Light’ approach (Birch, 2019). In short, the Theory-Light approach holds that we need not make major theoretical commitments in advance about the underlying mechanisms of consciousness. Instead, it suggests that in the first instance we identify plausible behavioural *markers* of consciousness via reference to human cases and then apply these to animals.

To give a simple example, let us suppose hypothetically that, on the basis of behavioural evidence, we conclude that humans cannot perform some task *T* involving stimulus *S* where *S* is presented subliminally. On the basis of this, if it were discovered that some non-human system NH was capable of performing task *T*, we might reasonably take this as evidence that the system was conscious, on the grounds that *T* could not be completed by a human without the involvement of conscious processing. This evidence would be further strengthened if we found that NH was capable of performing multiple distinct tasks like *T*, all of which required consciousness in humans.

An initial point to note is that the Theory-Light approach arguably captures the implicit methodology in much of cognitive ethology and the philosophy of animal minds. Researchers in this domain normally do not adopt an explicit theory of consciousness to begin with, but instead rely on a broad body of neurobiological and behavioural data to inform tentative judgements about which animals are likely to be conscious (Barron &

Klein, 2016; Bekoff, 2010; Birch, 2017; Griffin, 2001). Consider, for example, the following methodological statement from Griffin and Speck (2004):

> Although no single piece of evidence provides absolute proof of consciousness, [the] accumulation of strongly suggestive evidence increases significantly the likelihood that some animals experience at least simple conscious thoughts and feelings.

Of course, the kind of inference supported by the Theory-Light approach would be fallible and tentative; it might be the case, for example, that the mechanisms subserving task *T* in humans are different from those used by non-human system NH, such that performance of *T* by NH was not associated with consciousness. Likewise, it might be the case that some global necessary condition on consciousness (for example, a capacity for self-awareness) that is always present in humans is always absent in NHs. In that case, the ability to perform task T could be very good evidence of consciousness in humans and other systems in which the background conditions of conscious experience are realised, while providing very little evidence about consciousness in systems like NH.

Such considerations cannot be dismissed lightly, of course. Nonetheless, the hope of the Theory-Light approach is that via canvassing a large number of such markers, we might be able to make *reasonable inferences* about the likelihood of consciousness in different systems. Birch, for example, notes that at least three tasks that require consciousness in humans – namely, trace conditioning, reversal learning, and multisensory learning – can seemingly be accomplished by bees. In light of this clustering of markers for consciousness, he suggests that bees are excellent consciousness candidates, satisfying three key markers of consciousness in humans (Birch, 2019; see also Shea, 2012a).

The Theory-Light approach may seem to offer a straightforward improvement over the Theory-Heavy approach, allowing for reasonable assessments of consciousness in non-humans while deftly sidestepping the Specificity Problem. While I would agree with Birch that the Theory-Light approach is a valuable method for assessing consciousness in non-humans, it has several important limitations.

I take there to be three sorts of case where it may at best be uninformative and at worst be misleading if adopted too strictly. First, note that the Theory-Light Approach is somewhat lacking in resources for determining which of a system's mental states are conscious. In some cases it may be reasonable to infer that a given non-human's *perceptual* state is conscious, if, for example, it can serve as an input to a kind of learning or behaviour (such as trace conditioning) that we have reason to believe requires consciousness in humans. However, there will be a range of other states – emotional states, desires, and standing representational state such as beliefs, for example – that cannot be assessed in this manner.

In some instances, this may be of more than theoretical importance. Thus imagine we have reason to believe (perhaps on the basis of physiological indicators) that an organism is in a state of high stress, and wish to determine whether we have ethical grounds for administering a sedative. Even if we had reason to believe that the creature in question was capable of undergoing *some* conscious states, on the basis of its possessing various markers of consciousness, this would not allow us to automatically infer that its state of stress per se was associated with a conscious negative emotion. Nor is it immediately clear how a Theory-Light approach could settle this question, insofar as such emotions might not be able to serve as inputs to the kinds of intelligent behaviour we have identified as markers for consciousness.

A second limitation of the Theory Light approach concerns its ability to provide verdicts on systems that do not display the relevant markers of consciousness, and the risk of giving 'false negatives'. While capacities like trace conditioning when present may provide evidence of the presence of consciousness, it is not clear why we should take their absence as providing evidence of the absence of consciousness, especially when dealing with systems with quite different cognitive architecture from humans such as artificial systems or simpler animals. Consider, for example, that a creature heavily reliant upon just a single sensory modality such as smell might have little reason to acquire or retain a capacity for multisensory learning. Likewise, an organism that inhabited an environment with very stable environmental cues might gain little advantage from a capacity for reversal learning, despite being an intuitive plausible consciousness candidate for other reasons (for example, exhibiting sophisticated social intelligence in managing relations with its conspecifics). More broadly, I would draw attention to the fact that many conscious experiences in human beings such as the simple qualia involved in many perceptual experiences and bodily sensations seem at least prima facie unconnected to capacities for sophisticated learning, suggesting at least the possibility that consciousness may not require such capacities to begin with. In dealing with such cases where the markers of consciousness are absent, then, the advocate of the Theory-Light approach must either somewhat dogmatically rule out consciousness, or else remain agnostic, thus granting that the approach is limited in application.

A final limitation of the Theory Light approach is that there may be cases where the markers could be misleading, leading us to misattribute consciousness and give false positives. I have in mind in particular the possibility of artificial systems that were 'gerrymandered' specifically to exhibit our proposed markers of consciousness, while otherwise being fairly simplistic. Thus it seems at least prima facie possible that an engineer could build a robot capable of trace conditioning, reversal learning, and multisensory learning while otherwise being a very poor consciousness candidate (perhaps lacking capacities for any other forms of goal-directed behaviour, for example). While we cannot rule out that such a system would be conscious, it would surely be quite a different sort of case from that presented by a biological organism exhibiting those same markers, and thus requires special handling. What we would ideally wish in such a case – and what the Theory-Light approach seems unsuited to provide – is some principled way of assessing whether consciousness might be absent even *despite* the presence of the relevant markers.

6. A Modest Theoretical approach

I would suggest, then, that while the Theory-Light approach can avoid the Specificity Problem, it also has some limitations in its applicability, especially in the case of more phylogenetically remote organisms or more exotic systems. In this final part of the paper, I wish to propose a strategy for constructively combining Theory-Heavy and Theory-Light approaches in a way that allows us to overcome their respective limitations. Putting matters loosely to begin with, I would suggest that Theory-Heavy and Theory-Light approaches can operate in a form of *dynamic equilibrium*, with the insights of each informing and constraining the other. This is what I term the Modest Theoretical Approach.

There are several advantages of this strategy. First, and most importantly, it could provide us with some initial purchase on the Specificity Problem. Roughly, the idea is that we would use a markers of consciousness approach to identify species that are stronger and weaker consciousness candidates (cf., Birch, 2018). From here, we can examine the ways in which their cognitive architectures resemble and differ from those of humans. This evidence can then be combined with our preferred theory of consciousness so as to help us determine an appropriate level of specificity for our theory.

Applying this to GWT, for example, we might begin by first identifying a set of theory-independent markers of consciousness and use these to identify a set of non-human systems that are strong consciousness candidates. Let us suppose for the purposes of example that we then find that all these species (i) possess some form of brain-wide information-sharing, (ii) can selectively attend to sensory stimuli, and (iii) exhibit the "winner-takes-all" informational dynamics associated with global activation. However, we also find that some of these creatures lack episodic memory and domain general working memory. Such findings would give us at least some reason for thinking the appropriate level of specificity at which to spell out GWT for the purposes of comparative psychology is one that includes attention and winner-takes-all informational dynamics but does not make essential reference to episodic memory or domain general working memory. Having spelled out our theory in this way, we could then apply it to more difficult cases, such as systems that exhibit some but not all of the markers of consciousness already identified. If the systems in question nonetheless instantiated the cognitive mechanisms proposed by our theory (at the level of specificity identified in the preceding stage of inquiry), we might then have tentative reason to think they were conscious.

To give a concrete case where this method might apply, consider marine chordate lancelets (Fig.1). These organisms possess a lifecycle involving two distinct stages. In their larval stage, lancelets swim freely, but upon adulthood burrow into the seabed and remain largely stationary. Though lancelets possess sense organs (a single forward-facing eye), and a brain consisting of several thousand neurons, their neuroanatomy is significantly less complex than that of even the simplest vertebrates. Their range of behaviour is also correspondingly limited, and even during their more mobile larval stage lancelets engage in a few fairly simple behavioural 'routines' such as swimming upwards in the water column to feed or swimming quickly away from danger.

Fig.1. An adult lancelet.

Needless to say, there is little reason to expect lancelets to display sophisticated or versatile behaviours that might provide evidence of consciousness; they do not need to hunt food, for example, and thus need do little in the way of learning. Hence if we were to rely exclusively on a Theory-Light approach that used behavioural markers like trace conditioning to establish the presence of consciousness, we would likely have to remain agnostic about lancelet consciousness. By adopting the kind of Modest Theoretical Approach sketched above, however, we might be able to do somewhat better. Let us suppose for example that after careful testing, we find Lancelets exhibit only a handful of our markers of consciousness. We might nonetheless discover that Lancelets possessed some analogue of a Global Workspace, specifically those features identified as critical for consciousness at the preceding stage of inquiry. This might give us reason to consider Lancelets as potential or even reasonable consciousness candidates.[10] We might alternatively find that Lancelet neurobiology was too simple or too different to satisfy our theory of consciousness, thus giving us a negative answer. In either case, however, we would have arrived at a determination that – though fallible and tentative – would be difficult and perhaps impossible to reach just using a Theory-Light approach.[11]

[10] The idea that lancelets might possess an analogue of the global workspace may seem (and perhaps is) highly implausible, but it is worth noting that Dehaene himself claims that “consciousness is a useful device that is likely to have emerged a long time ago in evolution” (Dehaene, 2014: 244).

[11] This is simply an example, and I do not intend to make any substantive claims about Lancelet consciousness. Among researchers who have considered the question, most think that Lancelets are

The case of lancelets hopefully serves to illustrate how a Modest Theoretical Approach might help us to resolve the Specificity Problem and give verdicts about difficult cases. A similar method could be used to make determinations about the presence of consciousness in artificial systems. Thus we might find that an otherwise intelligent artificial system that did not exhibit any of our behavioural markers of consciousness nonetheless had an internal architecture relevantly like the human Global Workspace, thus making it a plausible consciousness candidate. Making this determination would only be possible, however, once we had some grounds for spelling out GWT at an appropriate level of specificity.

There are other advantages to a Modest Theoretical approach, including its ability to give clear verdicts about consciousness in unpromising cases like the gerrymandered robot discussed above. Specifically, once we have employed the markers of consciousness method advocated by the Theory-Light approach to flesh out our theory of consciousness at a given level of specificity, we may have principled reason in some cases to *disregard* the presence of the markers as a reliable indicator. Thus imagine that we find that all biological organisms that exhibit behavioural markers $M_1 \ldots M_n$ possessed a global workspace with features $C_1 \ldots C_n$. If we were then to encounter an artificial system that had been built specifically to exhibit markers $M_1 \ldots M_n$, yet which lacked all or most of features $C_1 \ldots C_n$, we might have good reason to doubt that the system was conscious. This would potentially allow us to exclude systems such as Searle's Chinese Room (Searle, 1980) that are intuitively weak consciousness candidates in spite of their superficial behavioural sophistication. After all, while the Chinese Room possesses numerous markers of consciousness (notably the ability to engage in verbal report), it will plausibly lack any analogue of the mechanisms identified as relevant to consciousness by the Modest Theoretical Approach, whether working memory, attention, or whole system information sharing. Of course, it might also be the case that it is impossible to build a system capable of exhibiting $M_1 \ldots M_n$ without also endowing it with features $C_1 \ldots C_n$. This would itself be a highly informative result, perhaps suggesting that the markers in

unlikely to be conscious; Feinberg and Mallatt, for example, claim that "most of the [lancelet's] sensory pathways have only three or four levels of neurons between the external stimulus and the motor response… [and] not the long processing hierarchy needed for sensory consciousness" (Feinberg & Mallatt, 2016: 48-9). Note, however, that they make this assertion only in light of their view that consciousness requires hierarchical sensory processing.

question were deeply connected with consciousness.

My discussion of the Modest Theoretical Approach thus far has focused on how we might use markers of consciousness to hone and flesh out a given theory of consciousness so as to make it applicable to tricky non-human cases. In closing, however, I would note two further useful applications unavailable to a Theory-Light approach alone. First, having spelled out our theory at an appropriate degree of specificity, we could investigate which of a creature's states are conscious, even those that were not revealed by markers to begin with. In principle, this would be a straightforward matter: having concluded, for example, that a creature's nervous system instantiated an appropriate analogue of the global workspace, we could determine whether a given mental state was broadcast on the workspace, or instead regulated behaviour in via non-centralised pathways. Answering such questions in practice would of course be far harder, but by integrating some measure of theory into a broader approach we can at least see a possible strategy for addressing them.

A second application for the Modest Theoretical Approach would be to use our markers of consciousness to help determine relatively more and less plausible theories of consciousness to begin with. Thus having used the markers to identify stronger and weaker consciousness candidates, we might assess whether a given theory can be coherently spelled out in such a way as to include the stronger candidates while excluding the weaker ones. If we discovered that there was no principled way of elaborating the theory that provided even a rough match with the determinations made via a markers of consciousness method, we might in turn come to have doubts about its plausibility. Putting matters crudely, if a theory cannot be spelled out in such a way as to include chimpanzees among conscious systems without also including pocket calculators, then this will be a defeasible reason to doubt the theory in question. This kind of determination can of course already be made on the basis of intuition. However, with recourse to markers of consciousness, we have a richer and more principled way of establishing strong and weak consciousness candidates without mere reliance on intuition, and thus will have a more nuanced test to which we can subject a given theory.

7. Conclusion

In this paper I have made three main claims. The first claim was that cognitive theories of consciousness face an important challenge in their application to non-human systems, namely the Specificity Problem. Second, I considered an alternative method for determining consciousness in non-human systems, namely Birch's Theory-Light approach, and suggested that despite its many virtues it had some important limitations. Finally, I claimed that the best strategy for resolving the Specificity Problem and making principled and accurate determinations of consciousness in non-humans was a Modest Theoretical Approach that integrates more traditional techniques from the theories of consciousness debate with the markers of conscious method proposed by the Theory-Light approach. I suggested that by using the markers of consciousness method we can determine an appropriate level of specificity to use when applying our theories of consciousness. Having done this, we can then extend our theories to explore possible cases of nonhuman consciousness, both biological and artificial, that would be hard to tackle using a Theory Light approach alone.

REFERENCES

Alkire, M. T., & Miller, J. (2005). General anesthesia and the neural correlates of consciousness. *Progress in Brain Research*, *150*, 229–244. https://doi.org/10.1016/S0079-6123(05)50017-7

Baars, B. J. (1988). *A Cognitive Theory of Consciousness*. Cambridge University Press.

Baddeley, A. (1992). Consciousness and working memory. *Consciousness and Cognition*, *1*(1), 3–6.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory?Trends in. *Cognitive Sciences*, *4*(11), 417–423.

Barron, A. B., & Klein, C. (2016). What insects can tell us about the origins of consciousness. *Proceedings of the National Academy of Sciences Proc Natl Acad Sci USA*, *113*, 4900–4908.

Bekoff, M. (2010). *The Emotional Lives of Animals: A Leading Scientist Explores Animal Joy, Sorrow, and Empathy — and Why They Matter*. New World Library.

Birch, J. (2017). Animal Sentience and the Precautionary Principle. *Animal Sentience*, *2*, 16(1).

Birch, J. (2018). Dimensions of Animal Sentience: Grain, Valence, Unity and Flow. *Unpublished*.

Birch, J. (2019). Invertebrate Consciousness: Three Approaches. *Unpublished Manuscript*.

Blake, R., & Logothetis, N. K. (2002). Visual competition. *Nature Reviews. Neuroscience*, *3*(1), 13–21. https://doi.org/10.1038/nrn701

Block, N. (1995). On a Confusion about a Function of Consciousness. *Behavioral and Brain Sciences*, *18*, 227–287.

Block, Ned. (1978). Troubles with Functionalism. *Minnesota Studies in the Philosophy of Science*, *9*, 261–325.

Block, Ned. (2007). Consciousness, Accessibility, and the Mesh between Psychology and Neuroscience. *Behavioral and Brain Sciences*, *30*(5), 481--548.

Block, Ned. (2009). Comparing the Major Theories of Consciousness. In M. Gazzaniga (Ed.), *The Cognitive Neurosciences IV* (pp. 1111–1123).

Boyd, R. N. (1999). Kinds, Complexity and Multiple Realization: Comments on Millikan's "Historical Kinds and the Special Sciences." *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, *95*(1/2), 67–98. JSTOR.

Boyle, A. (2019). *Mapping the Minds of Others*. https://doi.org/10.17863/CAM.37430

Brancucci, A., & Tommasi, L. (2011). "Binaural rivalry": Dichotic listening as a tool for the investigation of the neural correlate of consciousness. *Brain and Cognition*, *76*(2), 218–224. https://doi.org/10.1016/j.bandc.2011.02.007

Burgess, N., Maguire, E. A., & O'Keefe, J. (2002). The Human Hippocampus and Spatial and Episodic Memory. *Neuron*, *35*(4), 625–641. https://doi.org/10.1016/S0896-6273(02)00830-9

Camp, E. (2007). Thinking with maps. *Philosophical Perspectives*, *21*, 1–145.

Carruthers, P. (1999). Sympathy and subjectivity. *Australasian Journal of Philosophy*, *77*(4), 465–482.

Carruthers, P. (2005). *Consciousness: Essays From a Higher-Order Perspective*. Oxford University Press UK.

Carruthers, P. (2018a). Comparative psychology without consciousness. *Consciousness and Cognition*, *63*, 47–60.

Carruthers, P. (2018b). The problem of animal consciousness. *Proceedings and Addresses of the American Philosophical Association*, *92*.

Carruthers, P. (2020). *Human and Animal Minds: The Consciousness Questions Laid to Rest*. Oxford University Press.

Casali, A. G., Gosseries, O., Rosanova, M., Boly, M., Sarasso, S., Casali, K. R., Casarotto, S., Bruno, M.-A., Laureys, S., Tononi, G., & Massimini, M. (2013). A Theoretically Based Index of Consciousness Independent of Sensory Processing and Behavior. *Science Translational Medicine*, *5*(198), 198ra105-198ra105. https://doi.org/10.1126/scitranslmed.3006294

Crick, F., & Koch, C. (1990). Towards a neurobiological theory of consciousness. *Seminars in the Neurosciences*, *2*, 263–275.

Davidson, D. (1975). Thought and Talk. In S. Guttenplan (Ed.), *Mind and Language* (pp. 7–23). Oxford University Press.

Dawkins, M. S. (2012). *Why Animals Matter: Animal Consciousness, Animal Welfare, and Human Well-being*. OUP Oxford.

Dehaene, S. (2014). *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts*. Viking Press.

Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition*, *79*, 1–37.

Descartes, R., & Kenny, A. J. P. (1970). *Philosophical Letters*. Blackwell.

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

Graziano, M. S. A. (2013). *Consciousness and the Social Brain*. OUP USA.

Griffin, D. R. (2001). *Animal Minds: Beyond Cognition to Consciousness*. University of Chicago Press.

Griffin, D. R., & Speck, G. B. (2004). New evidence of animal consciousness. *Animal Cognition*, *7*(1), 5–18. https://doi.org/10.1007/s10071-003-0203-x

Kaminski, J., Call, J., & Tomasello, M. (2008). Chimpanzees know what others know, but not what they believe. *Cognition*, *109*(2), 224–234. https://doi.org/10.1016/j.cognition.2008.08.010

Kouider, S., & Dehaene, S. (2007). Levels of processing during non-conscious perception: A critical review of visual masking. *Philosophical Transactions of the Royal Society B*, *362*, 857–875.

Lamme, V. A. (2010). How neuroscience will change our view on consciousness. *Cognitive Neuroscience*, *13*, 204–220.

Lewis, D. (1972). Psychophysical and Theoretical Identifications. *Australasian Journal of Philosophy*, *50*(3), 249–258.

Lycan, W. G. (1996). *Consciousness and experience*. Mit Press.

Merker, B. (2005). The liabilities of mobility: A selection pressure for the transition to consciousness in animal evolution. *Consciousness And*, *14*(1), 89–114.

Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, *83*, 435–456.

Papineau, D. (2003). Could There Be A Science of Consciousness? *Philosophical Issues*, *13*(1), 205–220. https://doi.org/10.1111/1533-6077.00012

Prinz, J. (2012). *The Conscious Brain: How Attention Engenders Experience*. Oxford University Press.

Prinz, Jesse. (2018). Attention, working memory, and animal consciousness. In *The Routledge handbook of philosophy of animal minds* (pp. 185–195). Routledge/Taylor & Francis Group.

Regan, T. (2004). *The Case for Animal Rights*. University of California Press.

Rosenthal, D. (2005). *Consciousness and Mind*. Oxford University Press UK.

Rosenthal, D. (2019). Consciousness and confidence. *Neuropsychologia*, *128*, 255–265. https://doi.org/10.1016/j.neuropsychologia.2018.01.018

Rosenthal, D. M. (1986). Two Concepts of Consciousness. *Philosophical Studies*, *49*(May), 329–59.

Schwitzgebel, E. (2019). Is There Something It's Like to Be a Garden Snail? *Unpublished Manuscript*. https://faculty.ucr.edu/~eschwitz/SchwitzPapers/Snails-181025.pdf

Searle, J. R. (1980). Minds, Brains and Programs. *Behavioral and Brain Sciences*, *3*(3), 417–57.

Shea, N. (2012a). Methodological Encounters with the Phenomenal Kind. *Philosophy and Phenomenological Research*, *84*(2), 307–344. https://doi.org/10.1111/j.1933-1592.2010.00483.x

Shea, N. (2012b). Reward Prediction Error Signals are Meta-Representational. *Nous (Detroit, Mich.)*, *48*(2), 314–341. https://doi.org/10.1111/j.1468-0068.2012.00863.x

Simon, J. A. (2017). Vagueness and zombies: Why 'phenomenally conscious' has no borderline cases. *Philosophical Studies*, *174*(8), 2105–2123. https://doi.org/10.1007/s11098-016-0790-4

Singer, P. (1989). All Animals Are Equal. In T. Regan & P. Singer (Eds.), *Animal Rights and Human Obligations* (pp. 215--226). Oxford University Press.

Smith, M. R. (1994). *The Moral Problem*. Blackwell.

Stoerig, P., Zontanou, A., & Cowey, A. (2002). Aware or unaware: Assessment of cortical blindness in four men and a monkey. *Cerebral Cortex*, *12*(6), 565–574. https://doi.org/10.1093/cercor/12.6.565

Terrace, H. S., & Son, L. K. (2009). Comparative metacognition. *Current Opinion in Neurobiology*, *19*(1), 67–74. https://doi.org/10.1016/j.conb.2009.06.004

Thompson, L. T., Moyer, J. R., & Disterhoft, J. F. (1996). Trace eyeblink conditioning in rabbits demonstrates heterogeneity of learning ability both between and within age groups. *Neurobiology of Aging*, *17*(4), 619–629. https://doi.org/10.1016/0197-4580(96)00026-7

Tononi, G. (2008). Consciousness as Integrated Information: A Provisional Manifesto. *The Biological Bulletin*, *215*(3), 216.

Tononi, G., & Koch, C. (2015). Consciousness: Here, there and everywhere? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*.

Tsuchiya, N., Wilke, M., Frässle, S., & Lamme, V. A. (2015). No-Report Paradigms: Extracting the. In *True Neural Correlates of Consciousness. Trends in Cognitive Sciences* (Vol. 19, pp. 757–770).

Tye, M. (2017). *Tense Bees and Shell-shocked Crabs: Are Animals Conscious?* Oxford University Press.

Weiskrantz, L. (1986). *Blindsight: A Case Study and Implications*. Oxford University Press.