

Consciousness in Non-Human Animals: Problems and Progress

Peter Clark (S3316998)

Thesis for completion of BA in Philosophy of the Cognitive Sciences
Faculty of Philosophy of the University of Groningen
Under the Supervision of: Dr. Fred Keijzer
Additional Assessor: Prof. dr. Bart Streumer

June 18, 2021
Word Count: 9381

1. Introduction

The nature of consciousness has long been debated over the course of human history. Some of the world's greatest thinkers have made attempts at understanding and identifying what it is that makes humans conscious, and if that capability is present in any other organism or object. The status of attributing consciousness in non-human animals is presently unclear, with many avenues being pursued in order to clarify this. The situation becomes more complex when we consider the cases of phylogenetically diverse animal species such as fish, bees, octopi, and other invertebrates. Theories of biological consciousness need to provide answers to a few seminal questions about what is consciousness and its principal features, where and how it arose in evolutionary time in biological substrates, and does it exist as an all-or-nothing phenomenon or is its existence graded? These questions are derived from Browning & Veit (2021).

When studying consciousness in humans, research is directed towards identifying functional and anatomical markers of consciousness (Ginsburg & Jablonka (2019); Birch (2020); Shevlin (2021)), otherwise known as anthropogenic approaches. However, there are some issues with both the markers studied and their extrapolative power to non-humans. As Jonathan Birch has stated, "There is no agreement on whether any invertebrates are conscious and no agreement on a methodology that could settle the issue." (Birch, 2020). Ultimately, one of the issues for those working in the field is whether or not we can develop a theory of consciousness that minimizes or does-away-with the dependence on human markers of consciousness. Especially considering the many candidate animals that possess radically different neural anatomies. How avoidable is this anthropomorphic bias, and will an approach focused on how conscious experience arises help our efforts to demarcate consciousness in non-human animals?

In the study of cognition, Pamela Lyon (2006) proposed a biogenic approach that starts with biological cases and widens and defines the range of cognitive capacities that we find throughout the animal kingdom. More recently, Lyon and others referred to this as basal cognition (Lyon et al, 2021). They do so by introducing a toolkit of cognitive capacities found throughout the domains of life that can be studied independently in a wide range of animals. The toolkit includes functions like sensing/perception, memory, decision-making, learning, communication, valence and more. Their goal was to use these concepts from the domain of cognitive science and integrate it within the life sciences in a non-metaphorical way. As they state, "Arguably we will never fully understand the human mind until we understand its origins and evolution in the cognitive biology that [we] share with others in the living world." (Lyon et al., 2021). Seeing that consciousness is instantiated in biological beings, it is reasonable and desirable to develop a similar approach for a theory of consciousness, perhaps with a toolkit of consciousness-linked abilities. When I am referring to consciousness in this essay, I will be speaking of phenomenal consciousness, or phenomenally conscious experience (perception). This can also be referred to as subjective or felt experience. Additionally, this essay will focus on the case for biological consciousness.

In this thesis I will review modern approaches to non-human consciousness in the context of the problems inherent with measuring consciousness and defend the idea that phenomenal conscious experience comes from a biological toolkit of consciousness-linked abilities emerging from the underlying capacity for experience in any nervous system possessing organism. The modern approaches I will review include the Theory-Light and Modest Theoretical approaches from Jonathan Birch and Henry Shevlin, alongside the Gradualist Evolutionary approach from Peter Godfrey-Smith. Birch and Shevlin work on relaxing the

theoretical assumptions used in theories of human consciousness and posit that a robustly verified cluster of cognitive abilities will allow ascription of consciousness to non-humans. Godfrey-Smith starts from how the capacity for experience arises in the simplest forms of life, and that cognitive abilities that we relate to consciousness arise gradually over evolutionary time through a series of experience-shaping traits.

2. Measurement Problem of Consciousness

There is a measurement problem in the study of consciousness. The problem is how do we identify the features of consciousness and their properties? Consciousness at the moment is only directly accessible introspectively and in the study of it in humans, there is a reliance on verbal report, a capacity not present in animals (Browning & Veit, 2021). On top of that, verbal report is neither entirely accurate nor complete (Irvine, 2012; Ginsburg & Jablonka, 2019). When we consider candidates for identifying those features and their properties, the measurement problem gives us more to consider; the indicator validity and extrapolation problems alongside the importance of distinguishing unconscious processes and behaviors against non-conscious ones.

2.1 Indicator Validity Problem

When we are measuring something, we don't always directly measure the target, but instead use an indicator, marker, or indirect measure¹. The indicator validity part of the measurement problem has to do with the validity of these indicators when used to study consciousness. When using indirect measurements, there are a few inherent problems. For one, does the indicator actually measure the target state of consciousness? In many cases, what ends up happening is that you are only able to establish a correlation between the indirect measurement itself and target state of consciousness. If the target itself cannot be directly measured, as it seems to be the case of consciousness, validity cannot be established through these measures (Browning & Veit, 2021).

Another facet of this issue is understanding which distinction the indicator is making. Is the indicator meant to distinguish conscious vs non-conscious behavior, or is the indicator distinguishing between conscious vs unconscious behavior? Non-conscious in this case refers to cognitive processes and behaviors in an organism that does not possess consciousness nor the *scaffolding* for it. Unconscious refers to cognitive processes and behaviors in an individual that does possess consciousness, but the processes/behavior itself is not perceptible to conscious experience within that individual. This distinction is important, as we do not know whether it is the case that performing some behavior unconsciously equates to that same behavior performed non-consciously. Without knowing ahead of time, which studied animals are conscious or not, we do not possess any way to validate the indicators being used. To show that an indicator would be valid, we would need to show that it is present in a conscious animal, and absent in a non-conscious animal. However, we find ourselves in our current position where we do not have a way of demarcating whether an animal is conscious or not.

¹ Birch and Shevlin both use the term *marker* in place of indicator.

Desiring to validate indicators then, we would have to rely on background assumptions that associate the indicator to the target state of consciousness (Schickore & Coko, 2013). Considering that most of the indicators currently used in the study of consciousness in humans are ones aimed at uncovering a conscious/unconscious distinction in behavior, it seems that the assumption present is that the distinction of conscious/unconscious processes in humans will be the same as the distinction between conscious and non-conscious processes in other animals. This assumption lacks sufficient justification (Browning & Viet, 2021).

Some much-utilized examples of uncovering a conscious/unconscious distinction include the cases of trace conditioning, rapid reversal learning, and masking paradigms (Shevlin, 2021; Birch, 2020); more about these in a later section. In these cases, whether or not the participant was consciously aware of a stimulus significantly affected their responses. However, there are limited conclusions we can draw from this information; while we see evidence of conscious experience, we cannot answer whether the behavior performed by humans necessarily needs conscious experience, or if that behavior can be performed by a non-conscious organism.

It is important to note that all of the experiments rely on verbal report of conscious experience by the participant for their results, and thus their conclusions must be taken in context of the following. A capability to report may not be representative of conscious experience in an animal. Since we are unable to know the contents of consciousness, there is no clear method to judge whether the subjects are accurately reporting their conscious contents (Irvine, 2012). Another point made by Ginsburg and Jablonka (2019) is that there is an assumption for the link between report and conscious experience; it is possible that a subject has conscious experience of a stimulus, but fails to recall it correctly or lacks the ability to do so.

2.2 Unconscious / Non-Conscious Distinction

The cases of trace conditioning and rapid reversal learning, masking paradigms are set up to study the distinction between conscious and unconscious processes in humans, an animal that possesses consciousness. The maximum we can conclude is differences between conscious and unconscious processing in humans. There is no basis for the assumption that the distinction between conscious/unconscious processes should be the same distinction between conscious/non-conscious processes. It may be the case that some processes that require conscious experience in human subjects could be performed non-consciously by animals that evolved with different neuroanatomical or functional capacities. Second, there may exist processes capable of being performed unconsciously by humans, but not capable of being performed by non-conscious organisms due to them not possessing *the scaffolding of consciousness* (Browning & Veit, 2021).

Ginsburg and Jablonka (2019) give an example of the first point mentioned directly above. We can imagine as an individual develops consciousness, formerly non/unconscious processes move into the realm of conscious processing. Take the example of a young organism exhibiting innate unconscious food-seeking behavior. Initially the behavior is controlled through unconscious processes, but as the individual grows, learns from experience, and develops memories of those interactions, the behavior begins to come under

conscious control². Now we are in a position where the previous non/unconscious scaffolding can be shirked so that such food-seeking behaviors can now only be performed consciously. This doesn't mean that the process can't ever be done unconsciously again, just that in this case for this subject it cannot, as there are potentially many realizing mechanisms of complex behaviors. Put succinctly, the case that some processes are necessarily conscious does not extrapolate to them being necessarily conscious in all realizations.

Supporting the second point made by Browning and Veit above, it could be the case that there are unconscious behaviors that cannot be performed non-consciously due to the requiring the presence of consciousness at some point. These may include complex behaviors that require a learning phase. The neuroanatomy and cognitive processes that instantiate consciousness would be necessary as a framework for which these unconscious processes are then possible. This would mean that non-conscious organisms would not be capable of performing the aforementioned unconscious behavior. Take the example of the complex behavior driving home. Once you arrive at home, you notice that you have zero memory of the drive, as if the whole process was completed unconsciously. Ginsburg and Jablonka (2019) hold that while the complex behavior of driving may have been completed unconsciously, the consciousness scaffolding is still needed in the development and habituation of driving (and other complex behaviors). From this, it would mean that there are complex behaviors that can be performed unconsciously, but never non-consciously, therefore serving as potential evidence for consciousness.

2.3 Extrapolation Problem

Where the indicator validity problem focuses on whether indicators can differentiate conscious, unconscious, and non-conscious processes, the extrapolation problem focuses on whether indicators of conscious experience validated on human are useful and justified when studying conscious experience in non-humans. This problem arises in the case where we are able to justify the indicator choice when testing for consciousness in human subjects. The validated indicators of consciousness must be relevant to the mechanisms through which consciousness operates. At the time of writing this, it is not known if it is the same case for machines as for biological organisms.

When justifying human indicators of consciousness and extrapolating them to animals, we rely on the assumption that humans are *relevantly similar to animals*. (Browning & Veit, 2021) and that the processes and physical mechanisms between the measured indicator and conscious experience are likely to be of the same type. Justification for that assumption would have to come from anatomical and evolutionary evidence which is again, not extrapolatable to machines. For example, Berns (2018) while neuroimaging animals stated the following “analogous regions in dog and humans brains appear to serve analogous functions. This is important because analogous structure-function relationships provide a pathway to answering the question of what it's like to be a dog.”. Neuroanatomical similarities and analogous structure-function provide a potential for justifying extrapolation of human validated indicators of consciousness to evolutionarily close organisms.

² An example of this during an individual's development is breathing or blinking in humans. Both can be and usually are done unconsciously, but are capable of being pulled under conscious control by the subject.

Reliance on possession of the relevant biological similarities depends on how similar the organization, development and evolution of the animals in question are. As we move farther away from biological similarity, we become less and less justified in continuing to extrapolate with human validated markers of consciousness. That being said, there still exists some deep biological homologies at the developmental and cellular levels (Browning & Veit, 2021). Considering all justification comes from biological evidence, it is not useful to extrapolate in this manner to machines.

Using neural correlates of consciousness to extrapolate, even if we run with the idea that they are useful markers of consciousness, brings with it the risk that they may be too substrate and context specific to be useful for animals in different biological circumstances. Multiple realizability could solve the substrate and context dependence problem, but we still have the problem of validating indicators across those substrates.

At the present moment, presence of consciousness is a N=1 problem with the only verifiable sample arising in humans. Even if consciousness has arisen on multiple occasions, common ancestors in the evolutionary tree and the anatomical and developmental similarities that constrain evolution do not give us truly independent events (Browning and Veit, 2021). For example, the evolution of eyes initially thought to be a case of convergent evolution, was discovered to be a deep homology in the PAX6 gene. Perhaps this could be the case for certain consciousness-linked cognitive behaviors as well. Confirming the validity of indicators of consciousness and their extrapolation to non-human animals is necessary for any theory of consciousness.

3. Birch & Shevlin / Theory-Light and Modest Theoretical

Theories of consciousness try and provide sufficient conditions for conscious experience in humans. There are a few modern approaches to consciousness that vary based on the number of theoretical assumptions. Jonathan Birch (2021) labels these the theory-heavy, theory-neutral, and theory-light approaches.

Theory-heavy approaches rely on having a confirmed theory of consciousness in humans, and applying it as is to non-humans. The most prominent example of a theory-heavy approach is Global Workspace Theory (GWT) developed by Baars, Dehaene and others. The GWT posits that there is a global broadcasting mechanism in the brain that integrates representations from your perceptual, memory, and affective systems and broadcasts them into a *global workspace*, making those representations accessible to conscious experience. GWT posits that the global broadcast mechanism is located in the cortex, primarily the prefrontal cortex (Dehaene & Changeaux, 2011). There is strong empirical support for GWT, found in experimental paradigms including attentional blink and backwards masking in conjunction with neuroimaging techniques to differentiate processing in the brain resulting from conscious or unconscious perception of a stimulus (Dehaene, 2014). However, GWT requires that subjects be able to report their conscious experiences. This means it runs into fundamental issues when we consider the cases of humans with consciousness disorders that inhibit verbal report, and non-humans without the capacity for verbal report. All GWT can tell us is that the global broadcasting mechanism of a healthy brained adult human is sufficient for consciousness. All animals which we would share the indispensable features of the global workspace with, are conscious, and yet GWT can tell us nothing about which features are essential.

With some issues of theory-heavy approaches coming to light, Birch introduces the *dilemma of demandingness*, reminiscent of the extrapolation problem mentioned in chapter 2.3. The dilemma of demandingness states that “strong sufficient conditions for consciousness will not get us very far in making inferences about animals other than humans who can report their experience.” (Birch, 2020). While global workspace theory has strong empirical support, it is very limited in what it can tell us in non-human cases. There are other plausible theory-heavy approaches like Merker’s Midbrain Theory (MMT). MMT posits that the basis of human consciousness is not in the cortex, but rather in an evolutionarily older structure in the midbrain responsible for modelling the organism in its environment, pulling from many sources of sensory input and output to motor systems (Merker, 2007). The structure in humans is called the superior colliculus. MMT makes the case for consciousness in some insects. In insect brains, there is a functionally analogous mechanism called the central complex and it is believed by supporters of MMT that it too suffices for consciousness. While MMT allows for greater extrapolation to animals due to having less demanding sufficient conditions for consciousness, MMT has less empirical support. Verification of the sufficiency of the midbrain for conscious experience would require evidence of consciousness in humans who have lost the totality of their cortex, such as in cases of hydranencephaly. MMT however faces fundamental problems with the cases of blindsight and the fact that even in extreme cases of hydranencephaly, there is still some remnant cortical tissue and thus we cannot exclude cortical processing from consideration. If patient’s verbal reports are to be believed, then the case of blindsight would demonstrate that a functional midbrain is not sufficient for consciousness and instead conscious visual experience is dependent on the primary visual cortex (Doesburg & Ward, 2007).

With the over-reliance on verbal report and a set of strong sufficient conditions, theory-heavy approaches are restrictive in their application beyond fully functioning human adults. Theory-neutral approaches on the other hand elect to initially forego theoretical assumptions about cognitive and neural considerations. An example of a theory-light approach is the one posited by Michael Tye (2016), where he attempts to produce a collection of markers of pain in organisms, without committing to underlying theoretical assumptions. Tye uses the *same cause, same effect principle* when considering pain. Humans and animals behave in similar ways when presented with noxious stimuli, and we know that exposure to the noxious stimuli produces a feeling of pain in humans, so we are entitled to infer that a feeling of pain exists in non-humans as well. We are entitled to this belief unless we encounter a *defeater*, or negative indicator that would reject this notion. However, belief in the same cause, same effect principle is potentially credulous. What are the justifications that make two behaviors relevantly similar to each other? Learned avoidance behaviors for example, which often has a conscious experience of it in humans, has been observed in neural structure separate from the brain, such as in spinal cords of rats disconnected from the brain (Allen, Grau, & Meagher, 2009). It would be rash to use the same cause, same effect principle uncritically in these cases to say that conscious experience causes the learned avoidance behaviors. As Birch (2020) states, “considerations regarding cognitive and neural mechanisms may defeat the inference from surface behavior to conscious experience.”

How do we find what serves as something that would defeat that inference, what Birch calls a *defeater*? If we take certain neural mechanisms as defeaters, we would need to find humans who lack such neuroanatomical structures (cortex, etc.) but still possess conscious experience in order to show that said neural mechanism is not a defeater. Yet, this seems to negate the effort made by Tye to alleviate theoretical assumptions. Motivated by a lack of evidence regarding conscious experience in humans without a cortex,

a theory-neutral approach still is bound by a reliance on that evidence and other cognitive and neural considerations. Thus, we find that there is an *inescapability of theory*. Theory neutral approaches still need to demarcate what the negative markers and minimal system requirements of consciousness are, which leads us to the extrapolation issue with trying to use definite neural mechanisms as markers of conscious experience. One ends up returning to a theory-laden approach where we have to rely on (defeaters) minimum requirements for consciousness in order to say anything definitive about it.

3.1 Birch and the Theory-Light Approach

After entertaining the implications of theories of consciousness heavy or lacking in theoretical assumptions, Birch (2020) attempts to make progress on this problem by beginning with a hypothesis that makes minimal theoretical commitments, and is compatible with wide range of background theories. Birch's theory speaks of the relationship of phenomenal conscious experience and cognition; "phenomenal consciousness does *something* for cognition given the actual laws of nature, but precisely what it does is a question to which we do not yet have definite answers." (original emphasis, *ibid*).

The hypothesis proposed by Birch (2020) is termed the *facilitation hypothesis*. It states that phenomenally conscious perception of a stimulus facilitates, relative to unconscious perception, a cluster of cognitive abilities in relation to that stimulus. With this hypothesis, we are comparing processing in an organism happening when a stimulus is perceived consciously versus when it is perceived unconsciously. The main claim is that, *ceteris paribus*, conscious perception of a stimulus will facilitate a cluster of cognitive abilities. They are referred to as a cluster of cognitive abilities, because it is not intended that any ability alone serve as a marker or defeater of conscious experience. Instead there are "multiple consciousness-linked abilities which cluster together in the sense that there will be robust correlations between them." (*ibid*). The strongest case for consciousness as believed by Birch, is finding the set of consciousness-linked cognitive abilities that robustly co-vary across multiple timescales, ontogenically and phylogenically. This hypothesis is far from a completely specified theory of consciousness, and is *theory-light*. It is important to note that the distinctions studied by this hypothesis are still distinctions between conscious and unconscious processing of stimulus; what this means for how much Birch can speak about non-human consciousness will be explored later in the section.

The facilitation hypothesis speaks to why phenomenally conscious experience would be evolutionarily selected for, as it provides additional cognitive abilities to the organism. It is compatible with a variety of theories of consciousness, including ones that focus on the cortex (GWT) or midbrain (MMT), but rules out cognitive epiphenomenalism via masking trials. Evidence from human trials show that there is a cluster of cognitive abilities linked to conscious perception of a stimulus, such as with trace conditioning (Dehaene, 2014; Clarke et al, 2002), 2nd order conditioning (Ginsburg & Jablonka, 2019), rapid reversal learning (Travers et al, 2017), and cross-modal learning (Mudrik, Faivre, & Koch, 2014).

Trace conditioning for example, is a type of classical conditioning. The difference between trace and classical conditioning is when the second stimulus is presented, contemporaneously with classical, and with a temporal delay for trace conditioning. Humans can learn the association between the two stimuli, for example, a tone, followed a few seconds later by a puff of air aimed at your eye, only when they are consciously aware of both of the stimuli as well as the temporal interval between them. Participants were

however still able to complete classical conditioning and even delayed conditioning (classical conditioning but with overlapping stimulus presentations instead of contemporaneously) without conscious perception of the stimuli (Clarke et al, 2002). As soon as the temporal delay is added, conscious perception is required to learn the association. Trace conditioning results point to having conscious experience of a stimulus helps facilitate temporal cognition and learning temporal relations (Birch, 2020). It is important to remember that trace conditioning studies still rely on verbal report of the conscious experience of the participant.

Rapid reversal learning is when one learns the relationship between two stimuli, and has to learn the opposite relationship when reversed by the experimenter. In a study by Travers et al (2017), it was found that “subjects were only able to adapt [rapidly] to reversals of cue-target contingencies when consciously aware of the cues”. For this experiment, subjects were presented with priming arrows (‘<<’, ‘>>’) that were either unmasked or backwards masked. Regardless of masking, the arrows still primed a response direction. Subjects who were consciously perceiving the arrows were able to learn the relationship between priming arrow and target. When the association of stimuli was reversed by the experimenter, only those who were consciously perceiving the priming arrows were able to change their response patterns. Travers et al (2017) also found that a similar ability of judging cue reliability also followed the same pattern. When the cue became less reliable to the target, those who consciously perceived the change were able to accommodate it. We take this as evidence that having conscious experience of stimuli relationships *facilitates* fast associative learning when presented with unpredictable changes in the environment. For this experiment, experimenters relied on masking the stimuli to human perception.

It is through experimental paradigms, like stimulus-masking and manipulation, that we can study in humans which cognitive abilities are facilitated by conscious experience. The cluster of cognitive abilities is required, due to the fact that presence of a single ability in non-human animals doesn’t guarantee conscious perception is present, despite it being the case in humans. As we discover that more and more cognitive abilities from this cluster are potentially correlated, the case for conscious experience grows stronger. The list of cognitive abilities explored above is by no means exhaustive. Once we have this cluster of consciousness-linked abilities studied in humans, we can begin to search for abilities from that cluster in non-human animals, not only in individual species and the life of an individual, but looking whether the abilities switch on or off together across evolutionary time. A much relied on ability, verbal report, will notably be absent in non-humans.

Looking critically at Birch’s proposal, it is reasonable to say that we have shown that abilities are present, but not necessarily that they are necessarily facilitated by consciousness. Despite facilitation by conscious experience in humans, one can imagine that the abilities listed above could be possessed in non-humans without conscious experience. It may be that we are still only looking at indicators of the distinction between unconscious and conscious processing in humans. Since verbal report is not an available tool for studying consciousness in animals, the experimental paradigms have to focus on masking the perception of a stimulus, in ways like flash-suppression, backward masking, distraction, attentional blink and so on. The goal is to find a cognitive ability alongside a masking protocol to switch off that ability. Does switching off this ability also switch off other cognitive abilities in the consciousness-linked cluster, i.e. does a type of stimulus masking inhibit multiple cognitive abilities? If it were to be the case, the evidence would reciprocally support the claims that the abilities were indeed consciousness-involved and that the masking protocol successfully masked the desired stimulus. We would have an empirically supported hypothesis

concerning, in humans, the cluster of consciousness-linked cognitive abilities, evidence of that cluster in non-human target species with the same pattern of sensitivity to masking stimuli. As Birch (2020) states, we would then have “a scientifically credible inference to the best explanation of consciousness in the [studied species].”

Remembering the measurement problem as described in chapter 2, we run into a conceptual problem with the facilitation hypothesis. While Birch is calling for a systematic search for consciousness linked cognitive abilities, their relationship to each other, and their sensitivity to masking, it is still on the basis of evidence gained studying humans. There is a tradeoff between the evidence for sufficient conditions for consciousness in humans, and their relevance in extrapolation to non-humans. But again, Birch is not advocating to find sufficient conditions for consciousness. Instead, he uses the cluster as markers for presence of conscious experience in the same way symptoms provide evidence for the presence of a disease. For a disease, the presence of many related symptoms amounts to strong evidence of disease presence. Thinking this way, we can entertain the cases that a lack or presence of some symptoms doesn’t necessarily imply a lack of the disease. Likewise, presence of symptoms doesn’t necessarily entail presence of disease. This analogy can be applied to the cluster of cognitive abilities and attribution of consciousness.

Birch presents the case of bees as a candidate for possessing consciousness. When honeybees were presented with an odor, and then a sugar reward 5 seconds later, the bees learned an association (Szyszka et al, 2011). However, it is difficult to ensure immediate stimulus offset with odor, and it is thought that this evidence points to partial trace conditioning, where the temporal interval is not learned, but is a weakly conditioned response through association. There is evidence of bumblebees being able to learn temporal interval relations, an integral part of trace conditioning (Boisvert & Sherry, 2006). However, the study was done with a non-neutral stimulus (food reward), and so it is not trace conditioning in the same manner as mentioned above, due to the second stimulus being neutral. On the other hand, rapid reversal learning is a well-documented phenomenon in bees, but it is not agreed upon whether the sensitivity of bees to reversal increases or decreases over many reversals. The reversal experiment was set up with color associations to nectar (or no reward), at some point the colors are switched so the color associated with nectar now has no reward. For honeybees, their performance over reversals seems to deteriorate (Mota & Girufa, 2010), while in bumblebees it appears to improve (Sherry & Strang, 2014). While both types of bees displayed rapid reversal learning, honeybees seemed to learn the reversal less quickly as the experiment went on. Rats, goldfish, octopuses, and pigeons all demonstrate increased performance over successive reversals (Shettleworth, 2010) (Bublitz et al, 2017). Lastly, there is potential evidence of cross-modal learning in bees. After learning an association between a reward and a spatial scent pattern, bees showed a preference for visual arrays in the same layout as the scent pattern (Lawson et al, 2018). The bees seemed to be recognizing the similarities between their olfactory and visual patterns. However, it is still inconclusive to what extent cross-modal learning is facilitated by conscious perception in humans (Birch, 2020).

The benefit of looking for a cluster of cognitive abilities exhibited through behavior is that we are not reliant on theoretical commitments to neural architecture and organization. However, it might be the case that those animals with radically different neural architectures may only display a few abilities in the cluster, limiting extrapolative capability of human-centric markers. It may be that pursuing this approach can only give us weak evidence for candidate animals far removed from us in the evolutionary timeline. In its current state, Birch’s theory-light approach holds the potential for discovering consciousness linked abilities in

phylogenetically close organisms, but still suffers from the two main problems in measuring consciousness: a) masking stimuli to test for conscious perception is only testing the conscious/unconscious distinction and is insufficient for distinguishing conscious versus non-conscious experience and b) there is little extrapolative power to phylogenetically distant animals.

3.2 Shevlin and the Modest Theoretical Approach

Seeing the problems that Birch's *theory-light* approach was running into, Henry Shevlin (2021) proposes a modified version, which he terms *modest theoretical approach*. As we find more biological and even artificial candidates for consciousness, it is not likely that they will implement human neuroanatomical or functional architectures precisely, and so we run into what Shevlin calls the *specificity problem*. What is the appropriate level of detail that should be adopted by theories of consciousness when applying them to non-humans? In other words, how do we define the cognitive abilities and processes that constitute what we call consciousness, in such a way to make them extrapolatable to non-humans? If we are too specific, we run the risk of false negatives, and likewise, if the definitions for these cognitive abilities are too broad, we run the risk of misclassifying non-conscious systems as conscious. An approach to studying consciousness should answer which cognitive capacities are critical to consciousness, and how far from human architecture can a system diverge while still being conscious?

Take metacognition, usually understood as the capacity to internally represent one's own mental states. Metacognition can come in a range of different forms. It could involve something as simple as differentiating externally and internally generated changes to perceptual input (Merker, 2005), or something as complex as knowing that others can hold different beliefs (Kaminski, Call & Tomasello, 2008). However, the cognitive abilities and resources used in these metacognitive representations are unlikely to be similar in non-humans. Our language capabilities and conceptual understanding lets humans represent their thoughts as having a complex logical structure (Shevlin, 2021). We do not know yet if such a complex representation is required for non-humans to possess consciousness-relevant metacognition and higher-order thoughts.

It is reasonable to believe there is some level of specificity that allows us to make accurate attributions of consciousness to non-humans. So how can we establish principled criteria for assessing similarity across different non-human systems? These comparisons can still only be made on the basis of some further theoretical perspective of the relative importance of different cognitive systems. Birch's theory-light approach makes reasonable inferences about likelihood of consciousness in systems based on a number of markers like rapid reversal learning. As we elaborated on in the chapter above, there is still a large reliance on verbal report and we do not know if the relevant markers of conscious experience are sufficient or necessary conditions for consciousness. If the markers are present and we still cannot decide, then those markers are not sufficient indicators of consciousness and so we need a structured method of deciding whether consciousness might be absent despite presence of one or many of the relevant markers.

Shevlin brings up two other difficulties with the theory-light approach. First, Birch's approach lacks the resources to determine which of a system's mental states are conscious. It can allow us to infer that a non-humans perceptual state is conscious, on the basis of it being a source of input for a behavior that requires consciousness in humans (e.g. trace conditioning), but cannot tell us anything about the range of other

mental states we possess such as emotional states, or representational states like beliefs or desires (Shevlin, 2021). Those states cannot be studied by the theory light approach in its current form. The other difficulty Shevlin recognizes is the risk of false negatives. What can we say about organisms lacking the relevant markers of conscious experience? It is not clear why we should take the absence of any of the markers are sufficient for absence of conscious experience. Taking the examples suggested by Birch, we can intuitively imagine an organism that is heavily reliant on a single sense (e.g. olfactory), and therefore would have little reason to select for multi-modal learning. Likewise, an organism that exists in a very stable environment may have little reason to acquire rapid-reversal learning.

Considering the difficulties that arise in the theory-light approach, Shevlin presents his *modest theoretical approach*. The idea behind it is that we use an approach based on using markers of consciousness to identify potential candidate species for consciousness. In those candidates we then would compare and analyze the similarities and differences between their cognitive architectures and that of humans. We take the results of that comparison and then apply our desired theory of consciousness (e.g. GWT) to decide on a useful level of specificity.

Shevlin (2021) gives an example of how using this theory would look in the case of marine chordate lancelets and using global workspace theory. Lancelets are fish-like creatures with a single-forward facing eye and a brain containing a few thousand neurons. Following GWT, suppose we find that all Lancelets have a) a form of brain-wide information-sharing, b) selective attention to stimuli and c) demonstrates *winner-takes-all* information dynamics associated with global activation. We would then be able to evaluate candidate species at this level of specificity, making no reference to those types of memory. If an organism demonstrated some or all of the markers mentioned above, then we'd have reason to believe that the organism was conscious. Relying exclusively on behavioral markers like trace conditioning, we cannot make an ascription in the case of lancelets. Following the *modest theoretical approach*, we may discover that lancelets possessed some of the markers of consciousness as elaborated by GWT, or we may discover that their neurobiology was too simple. Either way, we have the ability through this approach to make a determination about the presence of consciousness in lancelets.

This approach would allow us to disregard some *gerrymandered* robots as conscious. Say that we discover that all biological systems displaying markers of consciousness ($M_1 \dots M_N$) possessed a global workspace with attributes ($C_1 \dots C_N$). If an artificial system displaying markers ($M_1 \dots M_N$), but doesn't possess the features ($C_1 \dots C_N$), we can reasonably exclude them from being conscious. This would apply to systems like Searle's Chinese Room.

Lastly, the modest theoretical approach could be used a judge of the different theories of consciousness. Once we have used markers of consciousness to identify weaker or stronger candidate organisms for consciousness, we can then evaluate whether any theory of consciousness agrees with presenting the strong candidates as conscious and the weaker candidates as potentially lacking.

Shevlin in his response to Birch has highlighted many places where theory-light approach alone falls short, and how allowing for a dynamic between theoretical commitments and the theory-light approach can advance the study of consciousness. Yet, have we managed to escaped the human bias in the selection of the markers of consciousness? While the modest theoretical approach gives us a more justified basis of

ascribing consciousness to non-humans, the basis of those markers still comes from study of consciousness in humans. Even if we were to find that presence of human markers of consciousness in a strong, phylogenetically close candidate organism may be enough to ascribe consciousness to that organism, judging on the basis of cognitive architectures still presents the human bias, and limits the extrapolation power of ascriptions of consciousness to those with functionally analogous structures as humans. In the following chapter, I will explore this problem from a different starting point: how subjective experience arises in biological organisms, and what that means for theories of consciousness.

4. Godfrey-Smith / Gradualism

Turning the evolutionary clock backwards, Peter Godfrey-Smith (2021) lays out the framework for how subjective experience arises in biological organisms, and how conscious experience may arise and exist in a gradual manner. The intent of his view is not to provide a framework to make ascriptions of consciousness to non-humans, but rather change how we think about the nature of subjective experience and provide support for a gradualist view of consciousness. Shevlin's specificity problem rears its head in the fact that animals lacking vertebrate neural architectures demonstrate complex behaviors associated with consciousness in humans, and so theories or approaches focused on how vertebrates do things might not end up including the basis for experience in the first place.

The modern view of subjective experience is that it is not just cognition viewed from the inside, as is empirically informed that much cognitive processing occurs unconsciously. A large part of the standard operation of sensing and cognition occur without any experiential accompaniment (Dehaene, 2014; Carruthers, 2019). The focus of study of experience has been directed at the distinction between conscious and unconscious processing within the category of cognitive activity, and the search for some process, some feature that makes cognitive activity *conscious*. That process, potentially something like an attentional spotlight or working-memory, is beyond what we consider the basic cognitive toolkit (Lyon et al., 2021) for an animal (Godfrey-Smith, 2021). It does not seem likely that later-evolving cognitive abilities would be the factor that *turns on* conscious experience in an animal out of its absence.

Godfrey-Smith (2021, p.10) introduces the idea of an *experiential profile* for each animal. A typical experiential profile has *breadth plus focus*. For example, in humans it is multi-faceted and has breadth including perceptual, affective, energetic and other aspects of experience. A single aspect of experience is able to take a large amount of focus. In all animals, the breadth side is more basic to experience and has to do with subjectivity and being the locus for sensing and action, while the focused side comes later, transforming experience by the additional abilities (such as working memory or an attentional spotlight) specific to evolutionary lines, yielding what Godfrey-Smith calls a derived experiential trait. Derived experiential traits shape or transform the already extant experiential profile of an organism. Different organisms may have different derived experience-shaping abilities or traits.

We are given the example of the inattentive long-distance driver. A driver is taking a long trip, that is routine or monotonous. The driver is able to respond to things happening on the road, minor corrections and the like, all while daydreaming or having the focus of their attention elsewhere. The driver then appears to *come to* after some time, realizing they have driven some distance. Godfrey-Smith posits that the driver possesses an experiential profile at each moment filled with the items in their attentional spotlight, sensory

elements (attended to and not), a mood and energy level, and so on. If the driver were attending to the same *daydream* or other thought, but not driving, then it would feel different to be them. Aspects of one's experiential profile may take the form of nested sets: the experiential profile, what is reportable, and what is in attentional spotlight. Figure 1 shows a graphical interpretation of the nested nature of experiential profile and the higher-level reportability capacities. The attentional spotlight can move around within the experiential profile, but there is a limit on what is reportable. The figure below is not meant to be exhaustive, but a conceptual example.

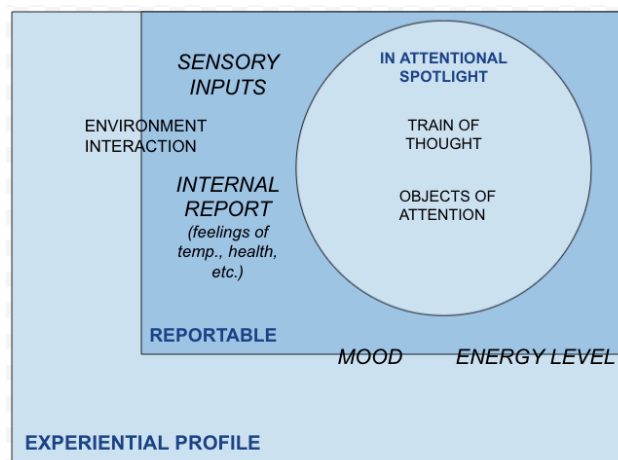


Fig. 1. Experiential Profile of Humans

This way of thinking could suggest that the integration of parts of the experiential profile serves as a physical explanation of consciousness. However, modern neuroscience has shown evidence that integration is not ubiquitous in complex animals. There are organisms displaying extensive lateralization, distinct sensory streams and other disunified complex cognitive features (Godfrey-Smith, 2021). Animals may handle the bringing-together of features of their experiential profiles through use of different derived traits. Integration may contribute to unity of conscious experience, but doesn't appear to be necessary. How does the idea of an experiential profile fit in to the evolutionary story then?

The evolution of experience happened naturally in the early stages of animal evolution (Godfrey-Smith, 2021). Basal cognition (Lyon et al., 2021) is nearly ubiquitous across living organisms, even down to all known cellular life. It consists of a toolkit of capacities including sensing and responding to the environment, some form of memory, and more. The toolkit of basal cognition contains subjectivity related features. These features were reinvented and manipulated in the evolution of multicellular organisms. Early nervous systems were focused on the coordination of action (Godfrey-Smith, 2020), especially the coordinated movement of body parts (Jékely et al. 2015). In studying a modern cnidarian, the polyp *Hydra*, it was found that they have four neural networks with distinct roles. These networks are not merely input-output networks, but rather they have been shown to produce oscillatory patterns that are affected by (albeit simple) sensory perception (Dupre & Yuska, 2017). While initially task-specific, we will see these nervous systems achieve some integration later in the evolutionary timeline.

A transition occurred with the evolution of bilaterally symmetric (*bilaterian*) bodies. They proliferated in the emerging predator-prey dynamics of the Cambrian Era. Those dynamics placed emphasis on

coordinated motion, and increased extero-sensing capacities, especially vision. Naturally, a change in nervous system organization accompanied this transition. These changes allowed animals to better handle refference; the effects of one's actions on the environment and actor's own senses. Neural circuitry dedicated to refference compensation has been found even in very simple bilaterians, a nematode *C. Elegans* with ~300 neurons. However, refferent circuits have not been detected in cnidarians yet. With this new nervous system organization, sensory processing is done in context of *self* vs *other*. Kaplan and Zimmer (2020) have demonstrated the presence of global patterns of brain activity associated with current behavior. These patterns can be affected on the global level by incoming stimuli.

Ancient nervous systems served a purpose of shaping action through patterns of neural activity that is modulated by the senses. These large-scale patterns are the platform for integration of sensory inputs, the differentiation of self vs other, are the basis for a stronger form of agency. These large-scale patterns can be affected by both internal and external stimuli, increasing in complexity with the development of refferent perception. Ginsburg and Jablonka (2019) show that if neural activity is organized in the form of large-scale patterns modulated by the senses, then a range of senses can affect any state. This whole process is the physical basis for subjectivity, and the basis for "an increasing *experientialization* of neural activity" (Godfrey-Smith, 2021). These patterns, while a system of control for behavior, also have a mapping to an organism's experiential profile.

With this, Godfrey-Smith hopes to establish that the scaffolding for phenomenally conscious experience is present in all nervous-system-possessing animals by means of large-scale dynamic neural processes affectable by internal and external stimuli, and that it arose gradually. Subjective experience does not have any particular function, but instead is the consequence of the evolution of action and coordination. Now we can address how these experiential profiles bear relation to conscious experience. What part of an organisms experiential profile is consciously accessible to them? An organism's experiential profile is heavily dependent on their sensory modalities, and may be radically different from our own. Each organism will have its own array of experience-relevant features. While our conscious experience is feature-based but bound (in the sense of the binding problem), it may be the case that an organism's experience instead is a "*multi-faceted wash*" (Godfrey-Smith, 2021), and their experiential profile is not fully integrated.

The later-coming experience shaping traits such as working memory or an attentional spotlight are not the basis for experience, but instead work on top of an organism's experiential profile. Defining when and where phenomenal conscious experience arises out of the experiential profile may be similar to defining the concept of life. There are a cluster of capacities necessary to consider an organism alive, such as metabolism and reproduction. This allows us to consider borderline cases such as viruses, who do not have a metabolism, but reproduce. Conscious experience then is best thought of as a cluster of capacities that allows for borderline cases to be considered (possession of one or some of the cluster). This means that conscious experience has a graded existence, dependent on possession of a cluster of experience shaping traits that allow for phenomenal access to an organism's experiential profile. These experience shaping traits may have evolved later in the evolutionary timeline and thus be limited to a narrow branch of the phylogenetic tree, but the underlying basis for experience is present throughout the nervous system possessing animal kingdom.

Now that we have laid out the biological basis of experience, it is time to look towards the future and how we can systematically search for experience-shaping traits which bestow consciousness-linked cognitive abilities in animals.

5. An Experientially Grounded Approach

With the acknowledgment of the ubiquity of some form of experiential profile among nervous-system-possessing animals and the fact that higher-level functions such as working memory or an attentional spotlight are not the basis for experience, but work on top of the profile, we can begin our search. While initially seeming disparate, the ideas of Birch and Shevlin and the ideas of Godfrey-Smith are two ends of the same rope. Let us see what an approach looks like taking their ideas into consideration, and how it stands up to the limitations of theoretical assumptions, the indicator validity problem, and the extrapolation problem.

Reminding ourselves of Birch's facilitation hypothesis from chapter 3 and Shevlin's modest theoretical approach; phenomenally conscious experience of a stimuli facilitates a cluster of cognitive abilities compared to unconscious perception. Current methodologies, like those followed by Birch and Shevlin, study the presence of phenomenally conscious experience in humans through evidence of the additional cognitive abilities that are not present when a stimuli is perceived unconsciously, e.g. trace conditioning. Once we have these markers (additional cognitive abilities), we use them to identify candidate organisms and analyze similarities and differences between their and our cognitive architectures.

We can expand Birch and Shevlin's approaches to include the language and ideas of chapter 4; phenomenally conscious experience of an aspect of an organism's experiential profile facilitates a cluster of cognitive abilities compared to unconscious experience. Phenomenal conscious experience is provided by a set of experience shaping traits. We study presence of phenomenally conscious experience in humans when the use of experience-shaping traits provides additional cognitive abilities. This allows us to test for those additional abilities in animals likely to possess the same experience-shaping traits. On top of searching for the cluster of cognitive abilities facilitated by those traits, we must extend our search to what experience-shaping traits are, and where and how they came about in evolutionary history.

Candidates for possession of those types of traits are sought out in two ways, synchronically and diachronically. Diachronically, we will look through the evolutionary tree and make educated guesses about when a certain experience-shaping trait would be likely to emerge. These traits emerge gradually, conceptually similar to Godfrey-Smith's account of nervous systems gradually arising to form the basis of experience. There is also no implicit assumption that the traits will emerge in the same gradation or fashion in different species, especially when there are discrepancies in sensory modalities. Synchronically, we can analyze and compare different organisms, who by behavioral analysis, functional and anatomical analysis, and evolutionary lineage, are likely to possess some variety of experience shaping traits that beget additional cognitive abilities. Lastly, we must attempt to identify experience-shaping traits that are either of a different nature to the ones we possess, or of a different gradation. In radically different neural

architectures than humans, such as in cephalopods, something like an attentional spotlight may manifest itself in a different manner, potentially disunified or even as multiple lower grade spotlights.

The benefits of extending our search for experience-shaping traits include the possibility (and in my opinion, strong likelihood) of discovering gradations of those traits in a diverse set of organisms, with each trait bestowing one or more additional cognitive abilities. Once we have a plausible candidate organism for possessing said traits, for example, an attentional spotlight, we can direct our search at studying the type and number of additional cognitive capacities that organism would have. Let us return to the case of two types of bees mentioned by Birch in chapter 3, honeybees (*apis mellifera*) and bumblebees (*Bombus terrestris*).

In the study of these types of bees, experimenters used a visual search task with distractors. The visual search task is often used in studying human psychology. In this task, bees were trained to seek a visual target of a certain color, in the presence of distractor targets of a different colors. It was found that while honeybees are *strongly influenced by number of distractors* (Morawetz & Spaethe, 2012), suggesting a serial-like visual search capacity, bumblebees are only *slightly affected* by distractor number, suggesting a more parallel-like search. One model proposed by Morawetz and Spaethe suggests that this difference can be explained by the fact that the attentional spotlight is larger in bumblebees than in honeybees, leading to the observed differences in visual search patterns. It is assumed that visual information processed in the attentional spotlight is done in parallel, not serially (Van Rullen et al. 2007). It is not known whether either type of bees can move their attentional spotlight around like humans can, but there is increasing evidence of flexibility of attention in them (Dyer, 2012). If this is the case that these two species of bees have slightly different attentional spotlights, then we have support for the idea that experience-shaping traits exist in different grades and forms in different species. Especially, because they are phylogenetically distant from humans, it lends merit to the idea that primitive versions of traits like an attentional spotlight evolved in a common ancestor between us and our buzzing friends, and we can expect to find forms of it in many an animal. The next step would be research how attentional control is instantiated in nervous systems.

Theories of consciousness are considered on top of the framework of experience given by an experiential approach. Let's look at a modified version of Tononi and Koch's (2015) Integrated Information Theory (IIT). As mentioned in chapter 4, while integration of information from sensory modalities is not necessary for complex behaviors, different grades of integration of experience could bequeath experience-shaping traits all the same. While IIT normally posits a minimal degree of integration in all things, biological or not, it would fit in with the experiential account by changing where the minimal degree of integration begins. The minimal degree of integration begins with an organism's unconscious experiential profile.

When studying conscious experience in non-humans, we removed a large amount of anthropocentric bias by first evaluating the scope of an organism's experiential profile to understand what kinds of information is accessible to the organism, and then searching for traits that can modulate and shape experience. When studying conscious experience in humans, the conclusions we can draw are about the type of conscious experience in an organism possessing the same or similar experiential profile. It is important to state that absence of additional cognitive abilities, or difference of apparent degree of experience-shaping traits do not serve as *defeaters* of conscious experience. Rather, they are an opportunity to discover where such traits or abilities began to emerge in the evolutionary timeline. Conceptually, the collection of experience-shaping

traits can be thought of as a toolkit for conscious experience. Possessing one of the traits in the toolkit, it can be used to facilitate a non-exclusive³ cluster of cognitive abilities.

Now we have a biologically grounded account for how experience arose from early nervous systems, and how conscious experience fits on top of that account through means of a series of traits that can shape experience, with the shaping of experience bestowing one or more of a cluster of cognitive abilities not present without said traits. The distinction between conscious and unconscious processes now exists on a graded spectrum and the distinction between conscious and non-conscious becomes trivial, any system without a nervous system would not possess the scaffolding for the biological type of conscious experience.

4

6. Conclusion

In this thesis the goal was to introduce the issues that come with studying consciousness in non-humans, analyze modern approaches with varying levels of theoretical commitments, explore the biological basis of experience in animals, and to defend the idea that phenomenal conscious experience emerges from a toolkit of experience shaping traits working on top of an organism's underlying experiential profile. This experientially grounded approach solves some of the modern issues that accompany studying consciousness, including the problem of justifying indicators of phenomenal conscious experience and extrapolating results to different animals. This approach allows for the study of consciousness and the search for experience shaping traits in many diverse and neuroanatomically different animal species than our own. The future of the study of consciousness using this approach may show that we are not alone in possessing phenomenal conscious experience, and in fact, we may be in good company.

³ Non-exclusive here is meant to reiterate that not possessing one or many of the cluster of cognitive abilities does not serve as a defeater of conscious experience.

⁴ This is not to say we are excluding artificial systems with dedicated sensory systems and something akin to large-scale patterns of activity that have reafferent responses, from being conscious, it is that we are exclusively looking at the case of biological consciousness.

Acknowledgements

Fred Keijzer for his unwavering support, guidance, and wealth of knowledge in the subject. Lukas Hummel for discussions, proof-reading and support.

References

Allen, C., Grau, J.W., & Meagher, M.W. (2009). The lower bounds of cognition: What do spinal cords reveal? In J. Bickle (Ed.), *The Oxford Handbook of Philosophy and Neuroscience* (pp. 129–142). New York: Oxford University Press.

Berns, G. (2018). *What It's Like to Be a Dog: And Other Adventures in Animal Neuroscience*. Oneworld Publications.

Birch, J. The search for invertebrate consciousness. *Noûs*. 2020; 1– 21.
<https://doi.org/10.1111/nous.12351>

Boisvert, M. J., & Sherry, D. F. (2006). Interval timing by an invertebrate, the bumble bee *Bombus impatiens*. *Current Biology*, 16, 1636–1640. <https://doi.org/10.1016/j.cub.2006.06.064>

Bublitz, A., Weinhold, S. R., Strobel, S., Dehnhardt, G., & Hanke, F. D. (2017). Reconsideration of serial visual reversal learning in octopus (*Octopus vulgaris*) from a methodological perspective. *Frontiers in Physiology*, 8, 54. <https://doi.org/10.3389/fphys.2017.00054>

Browning, H. & Veit, W. (forthcoming, preprint 2021). The Measurement Problem of Consciousness. *Philosophical Topics*.

Clark, R. E., Manns, J. R., & Squire, L. H. (2002). Classical conditioning, awareness, and brain systems. *Trends in Cognitive Sciences*, 12, 524–531. [https://doi.org/10.1016/s1364-6613\(02\)02041-7](https://doi.org/10.1016/s1364-6613(02)02041-7)

Carruthers, P. 2019. *Human and Animal Minds: The Consciousness Questions Laid to Rest*. Oxford: Oxford University Press.

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

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

Doesburg, S. M., & Ward, L.M. (2007). Corticothalamic necessity, qualia, and consciousness. *Behavioral and Brain Sciences*, 30, 90–91.

Dupre, C. and R. Yuste. 2017. "Non-overlapping Neural Networks in *Hydra vulgaris*." *Current Biology* 27(8): 1085-1097. doi: 10.1016/j.cub.2017.02.049.

Dyer, A. G. (2012). The mysterious cognitive abilities of bees: why models of visual processing need to consider experience and individual differences in animal performance. *J. Exp. Biol.* 215, 387-395

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

Godfrey-Smith, P. (2020). *Metazoa: Animal Life and the Birth of the Mind*. New York: Farrar, Straus and Giroux

Godfrey-Smith, P. (forthcoming, 2021) Gradualism and the Evolution of Experience. *Philosophical Topics*. [Link](#).

Irvine, E. (2012). Old problems with the new measures in the science of consciousness. *The British Journal for the Philosophy of Science*, 63(3), 627-648.

Jékely, G., F. Keijzer, and P. Godfrey-Smith. (2015). "An Option Space for Early Neural Evolution." *Philosophical Transactions of the Royal Society B*, 370: 20150181

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>

Kaplan, H. and M. Zimmer. 2020. "Brain-Wide Representations of Ongoing Behavior: A Universal Principle?" *Current Opinion in Neurobiology* 64: 60-69.

Lawson, D. A., Chittka, L., Whitney, H. M., & Rands, S. A. (2018). Bumblebees distinguish floral scent patterns, and can transfer these to corresponding visual patterns. *Proceedings of the Royal Society B*, 285, 20180661. <https://doi.org/10.1098/rspb.2018.0661>

Lyon P. (2006). The biogenic approach to cognition. *Cog. Proc.* 7, 11–29. (doi:10.1007/s10339-005-0016-8)

Lyon, P., Keijzer F., Arendt D., Levin M. (2021). Reframing Cognition: Getting down to biological basics. *Phil. Trans. R. Soc. B.* **376**: 2019750. <https://doi.org/10.1098/rstb.2019.0750>

Morawetz, Linde & Spaethe, Johannes. (2012). Visual attention in a complex search task differs between honeybees and bumblebees. *Journal of Experimental Biology*. 215. 2515-23. 10.1242/jeb.066399.

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

Merker, B. H. (2007). Consciousness without a cerebral cortex: A challenge for neuroscience and medicine. *Behavioral and Brain Sciences*, 30, 63–81. [https://doi.org/10.1016/S1053-8100\(03\)00002-3](https://doi.org/10.1016/S1053-8100(03)00002-3)

Mota, T., & Giurfa, M. (2010). Multiple reversal olfactory learning in honeybees. *Frontiers in Behavioral Neuroscience*, 4, 48. <https://doi.org/10.3389/fnbeh.2010.00048>

Mudrik, L., Faivre, N., & Koch, C. (2014). Information integration without awareness. *Trends in Cognitive Sciences*, 18, 488–496. <https://doi.org/10.1016/j.tics.2014.04.009>

Schickore, J. and Coko, K. (2013). Using multiple means of determination. *International Studies in the Philosophy of Science*. 27(3). 295-313.

Sherry, David F., & Strang, Caroline G. (2015). Contrasting styles in cognition and behaviour in bumblebees and honeybees. *Behavioural Processes*, 117, 59–69.
<https://doi.org/10.1016/j.beproc.2014.09.005>

Shettleworth, Sara J. (2010). *Cognition Evolution and Behavior* (2nd edition). Oxford: Oxford University Press.

Szyszkka, P., Demmler, C., Oemisch, M., Sommer, L., Biergans, S., Birnbach, B., ... Galizia, C. G. (2011). Mind the gap: Olfactory trace conditioning in honeybees. *The Journal of Neuroscience*, 31, 7229–7239.

Shevlin, H. Non-human consciousness and the specificity problem: A modest theoretical proposal. *Mind & Language*. 2021; 36: 297– 314. <https://doi.org/10.1111/mila.12338>

Tononi G, Koch C. 2015. "Consciousness: Here, There and Everywhere?" *Philosophical Transactions of the Royal Society* 370: 20140167. <http://dx.doi.org/10.1098/rstb.2014.0167>

Travers, E., Frith, C. D., & Shea, N. (2018). Learning rapidly about the relevance of visual cues requires conscious awareness. *Quarterly Journal of Experimental Psychology*, 71, 1698–1713.
<https://doi.org/10.1080/17470218.2017.1373834>

Tye, M. (2016). *Tense Bees and Shell-Shocked Crabs: Are Animals Conscious?* New York: Oxford University Press.

Van Rullen, R., Carlson, T. and Cavanagh, P. (2007). The blinking spotlight of attention. *Proc. Natl. Acad. Sci. USA* 104, 19204-19209