Can a marker approach exclude?

Jonah Branding[[1]](#footnote-1),[[2]](#footnote-2)

[brandin2@msu.edu](mailto:brandin2@msu.edu)

**Abstract**:

If an organism displays enough of the right neural, cognitive, or behavioral “markers,” researchers can generally assume it’s phenomenally conscious. But what if it doesn’t? Recently, there has been substantial disagreement on this “exclusion question.” According to one view (what I call the “symmetry” view), organisms lacking markers probably aren’t conscious (Dennett 1995; Tye 2016a; Birch 2022; Veidt 2022). However, according to another (the “asymmetry” view), we cannot conclude anything about the presence or absence of consciousness in organisms lacking markers (Prinz 2005; Schwitzgebel 2020; Andrews 2020, 2022, 2024). Here I argue that this disagreement partially reflects a deeper disagreement about how markers are identified; to this end, I point to three “paths” from specific ideas about marker identification (namely, *theory-based*, *analogy-based*, and *function-based* approaches) to one or the other view on exclusion. Equipped with the right auxiliary assumptions, theory- and analogy-based approaches can motivate the asymmetry view, whereas function-based approaches can motivate the symmetry view. However, this relation is not deterministic, as *different* auxiliary assumptions will lead to different views about exclusion. My distilled product is therefore a decision tree, which links views about marker identification to one or the other view on exclusion. I hope that this tree will serve as a means of identifying “fronts” where future debate on this question can be productively focused.

**Keywords**:

Animal minds, marker approach, consciousness, comparative psychology, ethology

**Statements and Declarations**:

The author has no competing interests to declare.

**Acknowledgements**

I thank Robyn Bluhm for generous and considered suggestions over the course of this project, plus Kristin Andrews, an anonymous reviewer, and audiences at meetings in New Orleans and Toronto. I also thank Elise Bisanz, who organized the New Orleans meeting, and Marriah Alcantara, along with Tyler Delmore, Will O’Shea, and Liz Waldberg, who together organized the Toronto meeting. Additional thanks to Tyler Delmore, for serving as a commentator.

“*The followers of the new reasoning [science] assert that the trees are sentient because they fold their leaves at night. [The traditional Buddhist response] was to say, “Well then, it must follow for you that pieces of leather are sentient because if they come near a fire, they shrink” … however, there are flowers named Santeu and Venus [flytrap] that, as soon as an insect lands on them, grab it, suck its blood, hollow out the body, and discard it on the ground … If such a flower were brought before us, would we dare contest their claim?*”

— Tibetan philosopher Gendun Chopel (1903-1951), qtd. in Lopez 2008, 123.

1. **Introduction**

For modern cognitive scientists as for Chopel’s “traditional Buddhists,” the Distribution Question—which asks which organisms are phenomenally conscious—is of significant moral and scientific importance. So phrased, it’s a metaphysical question; for short, call it DistributionM. Lurking in its shadow has always been another question, what you might call the “Exclusion Question” (or ExclusionM): which organisms are *not* conscious? Answering one of these questions suffices to answer the other: to get the non-conscious animals, just subtract out the conscious ones! However, there has recently been a shift in how the Distribution Question is conceived. Instead of straightforwardly asking which organisms are conscious, many authors (e.g. Birch 2022; Crump et al. 2022; Andrews 2022, 2024; Veit 2022) now ask which animals *we can know* are conscious. Since this is an epistemological question, call it DistributionE. Unlike in the metaphysical case, answering DistributionE does not suffice to answer ExclusionE, i.e., which organisms we can know are *not* conscious. It could be, for example, that we can know definitively when consciousness is *present*, but not when it’s *absent*. As I’ll argue here, there has of late been substantial, and rarely explicitly hashed-out, disagreement over precisely this issue.

The disagreement in question concerns “markers”: traits which are thought to indicate the presence of conscious experience, such as particular complex behaviors, patterns of brain activity or brain regions, or cognitive faculties. Organisms that display enough of the right kinds of markers are likely conscious. But what about organisms that display few or no markers? Lately, this question has received two different answers: (1) they are not conscious, and (2) we can’t say either way. These two answers reflect two different views about how we should interpret marker data.

According to the first view (call it the *symmetry* view), if good experimental evidence reveals that some organism O lacks enough of the relevant markers, we can conclude that O therefore lacks consciousness (Dennett 1996, 2019; Tye 1997; Mallat et al. 2021; Crump et al. 2022; Veit 2022; Solms 2022)—that is, for symmetry proponents, markers should be both *necessary* and *sufficient* for our making an attribution of consciousness. According to the other view (call it the *asymmetry* view), we simply cannot say anything about the presence or absence of consciousness in organisms who lack markers (Prinz 2005; Schwitzgebel 2020; Andrews 2018, 2020, 2022, 2024)—that is, for asymmetry proponents, markers should be *merely sufficient* for our making an attribution of consciousness.[[3]](#footnote-3) When they’re absent, we can’t conclude anything.

I believe the disagreement between these two ideas can—at least partially—be traced back to an even deeper disagreement: how should we go about identifying marker in the first place? For some, we discover markers by asking what a correct theory (or cluster of plausible theories [de Weerd 2024]) takes as necessary for consciousness. For those who accept this “theory-based” approach, concerns about multiple realization may motivate the asymmetry view (Prinz 2005; Schwitzgebel 2020), as our theories of consciousness may be inevitably “species-specific.” Others think we should identify markers through a process of analogy that begins with human intuitions about consciousness. Proponents of this “analogy-based” approach also have reason to endorse asymmetry, as human intuitions are possibly—or even likely—reliable for only an unrepresentative subset of consciousness-correlated features across taxa. For reasons we’ll discuss, this unrepresentativeness could bias a process of analogical reasoning that seeks to find consciousness wherever it occurs (Andrews 2018, 2020, 2022, 2024). Still others identify markers according to how well they illustrate that a particular *function* of consciousness is being executed (Birch 2022; Veit 2022). These authors may endorse the symmetry view, as we can observe a function’s absence just as well as its presence.

The central point of this paper is, therefore, the following. On two widely accepted approaches to the distribution question (theory-based and analogy-based approaches), there are reasons to think exclusion is not possible. On an alternative (function-based approach), there are good reasons to think it is. However, in no case is the link *determinative*. Rather, each of these logical paths have “off-ramps” for those who endorse a particular approach to justifying markers, but reject its associated implications for exclusion. For example, whether you think a theory-based approach permits exclusion turns out to depend on what you think about multiple realization; whether you think the same thing about a function-based approach depends on what you think about epistemological questions in the philosophy of evolutionary biology. Since most (all?) of these off-ramps begin with controversial philosophical or scientific issues, this paper won’t *answer* to the Exclusion Question; readers feeling lost in the woods on this issue will not be led out. However, they will get a map.

By a “map,” I mean a decision tree, linking *approaches to identifying markers* to *one or the other view on ExclusionE* (Fig. 5). I’m hoping my “map” will do two things. In the short term, I hope it’ll help *you* figure out where you should stand on ExclusionE, given several of your other philosophical views. In the long term, I hope it will identify “fronts” on which future debate over this questioncan be productively focused, or (less pugnaciously) provide a problem agenda for future research. In what follows, we’ll build the map in parts, beginning with the theory-based approach (Section 2), then looking at the analogy-based approach (Section 3), and finishing with the function-based approach (Section 4). Each section will return a piece of the map, which we’ll put together in the conclusion (Section 5).

1. **A theory-based path to asymmetry**

Let’s start with the basic issue: which animals are conscious? Perhaps the most obvious way to answer this question is to take what has been called a “theory-based” (Andrews 2020) or “theory-heavy” (Birch 2022) marker approach.[[4]](#footnote-4) This approach begins by developing a well-confirmed theory of consciousness in humans, and then asks, given this theory, which other organisms are conscious. This method is famously described by Daniel Dennett:

How, though, could we ever explore these “maybes”? We could first devise a theory that concentrated exclusively on human consciousness—the one variety about which we will brook no “maybes” or “probablys”—and then look and see which features of that account apply to which animals, and why” (Dennett 1995: 700).

At first glance, a theory-heavy approach seems like it might be able to deliver the symmetry view—i.e. it might permit us to say when consciousness is *absent*. Dennett himself thought it could when he claimed (for example) that if cephalopods moved “in the clunky way of most existing robots,” then it would be right to conclude that they are “some kind of zombies, marine robots with eight or ten appendages.” (Dennett 2019: 2) In this vein, Brian Key (2016) has argued that fish do not feel pain, because they lack the neural architecture responsible for pain in humans. However, there are also reasons to think that a theory-heavy approach can motivate the *asymmetry* view. To see how, let’s consider one theory in particular: Jesse Prinz’s (2005, 2012, 2018) “Attentive Intermediate Representations” theory (AIR).

According to AIR, consciousness in humans “arises when intermediate-level perception representations are made available to working memory via attention” (Prinz 2005: 388): a process Prinz takes to be underlied by patterns of phase-synched gamma oscillations in sensory neurons. Though the details of AIR’s empirical basis are mostly irrelevant to our purposes here, Prinz notes that its evidence comes largely from studies on human subjects.[[5]](#footnote-5) So far, this is in keeping with the theory-based methodology: we start with the human case, where we all agree that the organisms in question are conscious. However, when we try to *generalize* this theory, we encounter what Henry Shevlin has called the “specificity problem” (Shevlin 2021): is the product of our investigation a theory of consciousness *in general*, or a theory of consciousness *in humans*?

To flesh out this concern, suppose you weren’t interested in what makes organisms conscious, but in what enables bottle openers to uncork wine bottles. To study this, you take a bunch of double-wing bottle openers and observe what happens when they do their thing under various conditions. You then form the following theory: bottle-opening ability is caused by an upward pull from the screw mechanism on the cork; therefore, *screw* is among the “markers” of bottle-opening ability. Now suppose a dinner guest comes along with an air pressure bottle opener, which opens bottles by using a pump to pressurize the air beneath the cork. Consulting your theory, you judge the pressure opener incapable of opening your chardonnay, due to its lack of a screw. Taking bottle in hand, your guest proves you wrong, and pours you a glass.

What went wrong? In building your theory, you forgot that bottle openers provide a classic case of “multiple realization,” wherein some function, property, or process can have more than one causal or mereological mechanism (Putnam 1967). By looking only at double-winged openers, you’ve discovered necessary conditions for bottle-opening *within double-winged openers.* However, you have not discovered necessary conditions for bottle-opening *in general*, since there’s more than one way to open a bottle. Similarly, though having a certain array of properties (e.g. intermediate-level perception representations, gamma-wave oscillation, etc.) may be what *humans* need for consciousness, it would be premature to claim that these conditions are *necessary* for consciousness across taxa. Mixing metaphors, as Tim Bayne and colleagues have recently put it, “[a]rguably, trying to develop a comprehensive account of consciousness by studying only humans would be akin to trying to develop a comprehensive account of the elements by studying only copper” (Bayne et al. 2024: 455).

Though this quote might suggest an easy solution (just study consciousness in other animals!), such an effort is complicated by another issue, which I’ll call the “Distribution Catch-22.” In the bottle-opener case, there is a way to find out whether some entity has the property under consideration (bottle-opening ability): simply try to use it to open a bottle. This opens up the possibility for a *disjunctive theory* of bottle-opening, which could be accomplished in the following way: (1) Figure out what types of things open bottles, then (2) Sort them into types,[[6]](#footnote-6) and (3) Per the procedure with the double-winged opener above, figure out how each type works, until (4) You have a theory which: (a) states entity E can open bottles if and only if it’s a double-wing opener, a pressure opener, a wine key … and (b) describes the bottle-opening mechanism for each type in the classification system. The same schema, however, may be impossible for a theory of consciousness, since we lack the analogue of (1). Plausibly, we can only know whichanimals are conscious if we already have a general theory of consciousness; in order to form such a theory, we would need to know which animals are conscious![[7]](#footnote-7)

In this vein Eric Schwitzgebel has argued that theories of consciousness are inevitably “species-specific.” Schwitzgebel notes that though “[b]ehavior and physiology are directly observable… the presence or absence of consciousness must normally be inferred… once we move beyond the most familiar cases of intuitive consensus.” Since “[a]ll (or virtually all?) of our introspective and verbal evidence comes from a single species,” we get the following catch-22: (1) consciousness might be multiply realized, and (2) lacking the very theory of consciousness we’re after, we cannot determine the scope of its realizers so as to form a disjunctive theory (Schwitzgebel 2020: 57-58). Schwitzgebel illustrates this with the case of snail consciousness:

Let’s optimistically suppose that we learn that, in humans, consciousness involves X, Y, and Z physiological or functional features. Now, in snails we see X’, Y’, and Z’, or maybe W and Z”. Are X’, Y’, and Z’, or W and Z”, close enough? … In the human case, we might be able to – if things go very well! – rely on introspective reports to help ground a great theory of human consciousness… [but] without the help of snail introspections or verbal reports, it is unclear how we should generalize such findings (Schwitzgebel: 58). [[8]](#footnote-8)

Schwitzgebel holds that since consciousness might be realized differently in humans and snails, a theory of consciousness based on human evidence alone can only be a theory of *human* consciousness. Though *species*-specificity may be slightly too strict—a theory of consciousness developed for humans would apply to all creatures that are *relevantly similar* to humans, in that they instantiate the same “physiological or functional features” (here, X, Y, and Z)—it strikes me as correct in spirit. Plausibly, theories of consciousness are really theories of consciousness within relevantly physiologically and functionally homogeneous groups of organisms. We can investigate empirically whether some organism O falls into our group, by seeing whether it has X, Y, and Z; if O does, then on our theory, O is conscious. If O doesn’t, there’s nothing we can say, as our theory does not apply to O.

Prinz himself arrives at a similar conclusion, which he calls “level-headed mysterianism.”[[9]](#footnote-9) According to level-headed mysterianism, we will never have a complete theory of consciousness—of the kind that we would require to exclude—due to the Distribution Catch-22. As Kristin Andrews notes in her discussion, Prinz understands AIR as “a description of the neural correlates of consciousness in animals like us, not a theory about the nature of consciousness more generally” (Andrews 2020: 81). However, Prinz has also cited evidence that the category of “animals like us” is, given AIR, probably quite broad, including mammals, birds, and even insects, though possibly excluding cephalopods, fish, and reptiles. In these latter cases Prinz notes the apparent absence of AIR-consciousness may be due to a relative lack of documentation, but even if further work shows reptiles (say) are not “conscious-like-us,” they still may be conscious. Again, we simply cannot know (Prinz 2018).

This is all to say: level-headed mysterianism, or various related views such as “species-” or “group-specificity,” plausibly implies what I’m calling the asymmetry view. The presence of markers can be taken to provide positive evidence for consciousness in other organisms, by showing that those organisms are relevantly similar to humans. However, the absence of markers cannot rule out consciousness, as consciousness could be realized differently in these different systems. Due to the Distribution Catch-22, a disjunctive theory of consciousness cannot “cover” these various systems, as it isn’t clear, prior to a general theory, which organisms realize consciousness. This path to asymmetry is summarized in Figure (2a).

*Fig. 2a: The path to asymmetry, from theory-based approaches (arrows represent implication). If consciousness is multiply realized (1), then it may be implemented in types of organisms which our human-based theories don’t apply to—if so, then we get a catch-22: we must know who’s conscious to form a theory, and we must have a theory in order to know who’s conscious (2). Since a general theory of consciousness will thus remain out of reach (3), our species-specific theory will not allow us to make judgements about who isn’t conscious (4).*

But wait, that’s not all: our discussion so far also provides the symmetry proponent with a few “off-ramps,” which she could use to preserve her view of exclusion, even given a theory-based approach. Dennett and Key’s remarks suggest one way to do this: simply reject multiple realization! However, this isn’t her only way out. If she can find a way to around the Distribution Catch-22 (i.e. find a way to develop a disjunctive theory of consciousness), she could preserve the symmetry view, even while granting that consciousness might be multiply realized.

If you accept the theory-based approach, here’s your take-home message: to figure out which view of exclusion you ought to accept, simply follow this decision tree (Fig. 2b).

A white paper with black text

Description automatically generated

*Fig. 2b: The decision tree for the theory-based approach.*

In the rest of the paper, we’ll consider the same issue, for other approaches.

1. **An analogy-based path to asymmetry**

This section will consider how an analogy-based approach can also motivate asymmetry, using Kristin Andrews’ (2020) “Dynamic Marker Approach” (DMA). The DMA has been called a “theory-neutral” approach, because it does not assume a *theory* of consciousness (Birch 2022). Though it may seem strange that we could identify an organism as conscious without a theory, there are numerous cases in the history of science where something like this has been possible; long before we knew water was H2O, we were still able to identify water in most practical circumstances.

In constructing the DMA, Andrews begins by observing that our everyday attributions of consciousness are not mediated by *theories* of consciousness. This is certainly true in the human case. I do not infer that my roommate is conscious because she satisfies theoretical criteria; I simply recognize her as conscious due to the human way she acts and talks. “Non-inferentialism,” the idea that we can “directly perceive” consciousness in other humans and some animals due to an innate psychological faculty, is a view that goes back to Hume, to whom “no truth appear[ed] … more evident than that beasts are endowed with thought and reason as well as man” (Hume 2000, qtd. in Andrews 2020: 16). More recently, a non-inferentialist view has been defended by Dale Jameison (1998) and by Andrews herself:

“We see mind when having mind is the best explanation for what the individual does… Worries concerning whether there are other minds… are symptoms of a kind of philosophical sickness; they are just as fantastic as worries about the existence of the external world” (Andrews 2018: 244).

Like Dennett, Andrews suggests we begin in the human case, where we “brook no ‘maybes’ or ‘probablys’.” However, from there Andrews doesn’t require we form a general *theory* of consciousness. Rather, we can look to introspection and social psychology to figure out what observable features of other human beings cause us to perceive them as minded. Andrews calls these features “initial markers.”

We can then use this set of initial-marker organisms to identify “derived markers,” which are other features found to cluster with initial markers: for example, certain neurotransmitters, behaviors, or brain regions. The more derived and initial markers an organism has, the likelier it is that that organism is conscious. In practice, this might look like the following. Suppose we discover having human-like facial expressions is the only initial marker—in other words, we view a given organism as minded if and only if it has human-like facial expressions. Since only humans and other primates have the relevantly human-like expressions, our candidates for consciousness are {humans, other primates}. Then, we compare these organisms, and ask what other features they share in common. As it turns out, they both have prefrontal cortices (PFC’s), so we put this down as a derived marker. We then ask what *other* organisms have this derived marker. Suppose we find this is all mammals. We might then expand our list of candidates to {all mammals}, with higher credences for primates, due to their also satisfying the facial expression criterion.

This last note is important: the DMA does not provide a list of conscious organisms *per se*, but a dynamic list of *candidates* for consciousness, with associated credences. For example, suppose we find that platypi, though possessed of prefrontal cortices, have few other of the initial or derived markers. This would lead us to downgrade our credence in their consciousness. Andrews illustrates this nuance with the example of Large Language Models:

“Typically, if someone is able to talk, we take them to be conscious. However, suppose next year Google’s DeepMind creates a chatbot that thoroughly passes the Turing Test… [even then] without nonverbal expressions of pain or emotion—that is, without any other marker of consciousness—even humans who were first fooled by the bot may be disinclined to take the chatbot as conscious once they realize that it lacks all the other markers…The lesson is that language is one marker, but one marker alone is not sufficient for concluding that some being is conscious” (Andrews 2020: 87).

So, the DMA is an alternative or supplement to theory-based approaches, which delivers a dynamic list of candidates for consciousness using a set of initial markers. I take it that the DMA is currently the most fleshed-out analogy-based marker approach, though our remarks in what follows will apply to *any* analogy-based approach that relies on identifying a set of initial markers (e.g. Varner 2012).

Now let’s get to the interesting question: how do we get from an analogy-based approach to marker identification to one or the other view on exclusion? There is at least one pretty clear path, and it leads to the asymmetry view.

To see it, we might begin by considering how the DMA is relevantly similar to an unsupervised machine learning algorithm: instead of employing a specific theory of consciousness, and then testing whether some organism meets the criteria of that theory, it uses known data (that initial marker-organisms are conscious), to find relevant similarities (the derived markers), and then generate a prediction (that derived marker-organisms are also conscious). I make this analogy because, like unsupervised learning, the DMA is prone to a problem called *selection bias*, where an algorithm can make faulty predictions if its training data is non-representative. For example, if I’m trying to program an algorithm to detect whether an image is a flower, and I only train it with yellow daffodils, it may decide that a purple pansy is not a flower. To get the algorithm to correctly sort flowers from non-flowers, we would need to start with a representative sample of all flowers. Similarly, if we take only *organisms displaying initial markers* as our “input data,” our predictions may well be biased, because there’s no guarantee that initial marker-organisms are representative of *all* conscious organisms. To know whether the DMA could exclude, we would need to know whether our sample (initial marker-organisms) is representative: something we *can’t* know without priorinformation about the total set of conscious organisms.

This fact—that we can’t know whether the initial-marker set is representative—is itself enough to secure asymmetry, since after all, this is the idea that we cannot *decisively* *rule out* consciousness in unmarked taxa. However, there are even good reasons to think it’s *not* representative, given currently popular ideas about human Theory of Mind. For instance, suppose our ability to recognize initial markers is an innate modular faculty, which evolved to facilitate “perspective-taking,” itself useful in dealing with complex primate social hierarchies (Brüne & Brüne-Cohrs 2005). On this account, the initial markers evolved because they track *human*consciousness, and would therefore (plausibly) be highly *un*representative. Andrews has recognized this, having claimed, for example, that “[f]or individuals who move in ways different from us, because their bodies are morphologically very different from our own—fish, spiders, bees, whales—we may fail to pick out the meaningful patterns that reflect the mindedness of the species” (Andrews 2018: 242). This idea gives her good grounds to accept an asymmetry view of analogy-based approaches: “sentience tests [i.e. experimental paradigms designed to detect the presence of markers] … are not a decision procedure to sort sentient from non-sentient beings,” but instead offer “positive tests for sentience—new ways of seeing particular sorts of sentience” (Andrews 2022: 1-2).

To take yet another artifact analogy, it might be helpful to think of positive markers as like metal detectors.[[10]](#footnote-10) Metal detectors work by transmitting an electric current into the ground, which reflects off of high-conductivity objects like gold rings. This design implies a limitation: there are metals which a metal detector cannot detect, because some metals (such as stainless steel or titanium) aren’t highly conductive, and therefore won’t reflect the current. Generally this isn’t a problem, since metal detectors were not designed to find these types of metals. However, it does mean that while metal detectors can provide *positive* tests for the presence of metal, they cannot provide tests for the absence of metal in general, merely the absence of *detectable* metal. Similarly, marker evidence may only provide evidence against types of consciousness that are “reachable” through a process of analogical inference.

In summary, since analogy-based approaches like the DMA—which is ultimately justified by our innate ability to recognize certain kinds of consciousness—are liable to selection bias, they may well leave us with asymmetry (Fig. 3a).

Fig. 3a: *The path to asymmetry, from analogy-based approaches. If we begin with our intuitions about consciousness and proceed by a process of analogy (1), we may be limited if our intuitions are unrepresentative (2), and thus a process of reasoning with our intuitions as primary analogues would not be able to deliver a negative verdict on organisms who are conscious, but “differently marked” (3).*

Does the symmetry proponent have a way to resist this conclusion, given an analogy-based approach? The most obvious idea would be to reject the possibility of selection bias. She could accomplish this, for example, by arguing that our innate consciousness-recognizing module evolved to see minds *wherever they occur*, not just in humans. To me, this seems like a tall, and highly speculative, task. The better way to go may be to reject the 2-3 arrow of implication in Figure (2a): sure, our initial markers might be biased, though this doesn’t necessarily leave us with asymmetry. This could well be the case, if we believe there are “defeaters.”

A “defeater” (or “negative marker” [Andrews 2024]) is an observable feature of an organism that *undermines* an inference to consciousness. Following Michael Tye (2016b), suppose that after some guy Bob’s death, we discover Bob lacks a brain. Though throughout Bob’s life, his rich, human-like behaviors have led us to see him as conscious, you might think that this discovery should undermine that inference. In other words, Bob’s lacking a brain is a “defeater.” For Tye, defeaters aren’t only the stuff of science fiction; he thinks a real-life analogue is provided by the discovery that *C. elegans—*an organism capable of some quite sophisticated behavior—only has around 300 neurons (Tye 2016a, 200). The point is: just as positive markers indicate the *presence* of consciousness, negative markers—such as Bob’s lacking a brain and C. elegans’ lacking enough neurons—indicate its absence. In terms of our metal detector metaphor, this would be like if you knew that a certain plant would only grow in the absence of *any* buried metal. If you see the plant (= negative marker), you could know where there’s no metal (= consciousness), even though your metal detector (= positive marker) couldn’t tell you this on its own.

But then, the crucial question: *are there defeaters/negative markers*? Some think there aren’t—or at least, if there *are*, then they would defeat the purpose of adopting the analogy-based approach anyway. Jonathan Birch (2022), for example, asks how we could know whether some feature is, or is not, a defeater. Looking at several of Tye’s examples, he suggests that the underlying principle is: “to show that the absence of some neuroanatomical structure… is not a defeater, we need to find evidence in humans showing that brains without that structure still have conscious experience” (Birch 2022: 139, emphasis removed). However, if we know this, Birch claims, we already know enough to use a theory-based, instead of an analogy-based, approach! Andrews herself has made a similar point, claiming that negative markers require theory-laden assumptions that today’s science cannot supply:

[N]egative markers rest on assumptions about what properties are necessary for consciousness, at a point in the science when we do not even know what kind of neurobiology is sufficient for consciousness. For example, a commitment to a central nervous system as necessary for consciousness may be seen as coming from a commitment regarding the function of consciousness, namely that it is for binding and broadcasting signals, plus the commitment that a central nervous system is needed to achieve this function. These are two theory-laden assumptions (Andrews 2024: 425).

Here's what this means for you, uncommitted reader. If you accept the analogy-based approach, follow this tree to see where you fall on exclusion.

A diagram of a question

Description automatically generated with medium confidence

*Fig. 3b: Decision tree for the analogy-based approach.*

Before putting this all together, we have one last approach to consider: the function-based approach.

1. **A function-based path to the symmetry view.**

What I will call a “function-based” approach has elsewhere been called a “theory-light” approach (Birch 2022). The distinguishing feature of this approach is that, while it doesn’t endorse a *theory* of consciousness, it does make some claims about the *functional role* of consciousness. To take a theory-light approach is to be committed to what Jonathan Birch (2022) calls the “facilitation hypothesis” (FH):

FH: Phenomenally conscious perception of a stimulus facilitates, relative to unconscious perception, a

cluster of cognitive abilities in relation to that stimulus. (Birch 2022: 140)

For Birch, consciousness has a certain functional role, which consists in facilitating a suite of cognitive abilities. Organisms who have these abilities therefore are (or likely are) conscious. Organisms who don’t aren’t (or likely aren’t). Hence, this “function-based” approach can lead us to the symmetry view, which these sorts of authors often endorse. For example, Birch’s own group has claimed that even in an organism possessed of one “indicator” (= marker) of interest, “[i]f high quality scientific work shows the other indicators to be absent, then sentience is unlikely” (Crump et al. 2022: 2). Considering insect consciousness, Walter Veit acknowledges that though “lack of evidence [of consciousness in insects] … should not be confused with evidence of absence” (Veit 2022: 300), this is merely because the relevant work is not yet done; once we have more thorough studies, we can expect to deliver a decisive verdict on consciousness in insects.

The point is: if we can establish a general *function* of consciousness—consisting, for example, in facilitating some set of cognitive abilities—we can deliver negative, as well as positive, verdicts on distribution. But what *is* this function of consciousness? What specific abilities is consciousness supposed to facilitate? As a preliminary list, Birch (2022) suggests three: trace-conditioning, rapid-reversal learning, and cross-modal learning. Why these three? Because we know that in humans, these sorts of behaviors require conscious perception of stimuli—they’re not the kinds of things you can do, for example, while sleepwalking. Though the facilitation hypothesis in its broad sense does not require specifically *these* cognitive abilities (the idea is simply that there is *some* evolved function of consciousness, *F*, and we entertain one set of markers *M* just in case M reliably indicates that F is actually being fulfilled), the capacities endorsed in other recent accounts are broadly similar.

Mark Solms, for example, has suggested that the “fundamental basis of qualia” is to allow us to “treat our needs categorically” so as to enable decision-making among multiple competing needs. Considering the case of our felt urge to inhale when holding our breath, Solms asks: “What does the feeling of your blood gas balance add to their normal autonomic control? It adds to the capacity for *choice*, which underpins voluntary (as opposed to automatic) action” (Solms 2022: 2). According to Walter Veit, “[t]he function of consciousness is to enable an agent to respond to … the computational complexity of the economic trade-off problem between competing actions faced by all organisms as they deal with challenges and opportunities throughout their life cycle in order to maximize their fitness” (Veit 2022: 293). In other words, some organisms often find themselves in situations where they need to make trade-offs between different possible behaviors so as to maximize their fitness. Organisms who evolved the capacity to make these kinds of tradeoffs are the conscious ones; organisms who didn’t are the non-conscious ones.

The theme of enabling trade-offs and behavioral complexity runs through other accounts as well. In that same publication from Birch’s group, the authors suggest that sentience may underlie operant conditioning, because it “allow[s] animals to ‘label’ previously neutral stimuli, or behaviors with valanced information, so that the animal can learn to behave more beneficially when faced with similar situations in the future” (Crump et al 2022: 8), though they consider another “plausible evolutionary function of sentience” (Crump et al: 17) to be weighing different motivations in decision-making. In the same spirit, Peter Godfrey-Smith (2020) suggests that consciousness may have evolved along with the fast-moving dynamic body types of the Cambrian period (~589-485 mya). Why these body types in particular? Because since “Sensing has its *raison d’etre* in the control of action… The evolution of mind includes the coupled evolution of agency and subjectivity” (Godfrey-Smith 2020: 59), and thus “the evolution of animal agency brings with it the origin of subjects” (Godfrey-Smith: 105).

If I were to summarize the “function of consciousness” implicit in this family of statements, I’d say something like this: the function of consciousness is to integrate information from (possibly) multiple sensory modalities, both synchronically and diachronically, in order to make decisions about what to do in novel situations.[[11]](#footnote-11) Now, assuming this (or something like it) *is* the function of consciousness, we can delineate markers in accordance with how well they demonstrate that this function is being accomplished: when we see flexible decision-like behaviors informed by multiple sensory modalities, we should make attributions of consciousness. Since these attributions seemingly depend on neither the truth of any background theory of consciousness nor the reliability of human intuitions, symmetry is secured.

Here’s an example of how this reasoning might be applied in practice. Hermit crabs, unlike other crabs, don’t naturally grow a carapace, and instead must select shells from their environment. As the right shell to select is context-dependent, the crabs seemingly find themselves in a situation where flexible information processing of the sort provided by consciousness would be useful; and in fact, it appears the crabs *do* process information in this way. They select heavier shells in stronger currents, and when separated from their previous shell, will prefer shells they’ve already used in the past to novel weight-matched shells (Hahn 1998). Hermit crabs also display motivational trade-offs: they will forgo a higher-quality shell for a lower-quality shell if subjected to shocks in the higher-quality shell (Appel & Elwood 2009), though when shocked they’re slower to evacuate a high-quality than a low-quality shell (Elwood 2009). The crabs will also swap a higher- for a lower-quality shell if the lower-quality shell can fit through a narrow pass in an escape maze (Krieger, Hörnig, & Laidre 2020). [[12]](#footnote-12)

Meanwhile, organisms who don’t display this kind of flexible processing may be, in Dennett’s phrase, effectively “zombies” or “robots.” Feinberg & Mallatt (2016), using markers designed to show “that animals can recognize, experience, and remember rewards versus punishments,” (Feinberg & Mallatt 2016: 174) conclude that roundworms and flatworms lack consciousness, since (for example) they appear to lack behavioral tradeoffs, and won’t self-deliver analgesics or rewards (Feinberg & Mallatt, 175; though see Andrews [2022] for a different interpretation of the literature here); someone who thinks behavioral trading-off is a consciousness-facilitated capacity would agree that this is evidence against consciousness in roundworms. The lesson: a function-based approach to justifying marker data can seemingly do what some have deemed impossible: legislate negative answers to questions of animal sentience (Fig. 4a).

Fig. 4a: *The path to symmetry, from function-based approaches. We begin by identifying a function of consciousness, consisting in the set of capacities we know consciousness facilitates in humans (1). Then, we can set up experiments to test whether this function is executed in various organisms (2). If it isn’t, we can infer that the organism is not conscious (3).*

Now, (finally!) the burden is on the *asymmetry* proponent, to try and come up with *his* set of off-ramps.[[13]](#footnote-13)

One way out would be to challenge the facilitation hypothesis, by asking the following question: *why would we say that the function of consciousness is flexible decision-making, or whatever else it’s taken to be*? Solms suggests this is because we know from introspection that “executive control” comes along with consciousness. Birch suggests that since consciousness facilitates (e.g.) cross-modal learning in humans, it probably also facilitates them in other animals. More classical arguments for a “function” of consciousness have included weakened abilities of discrimination that go along with human blindsight and sleepwalking (Block 1995). However, there are a couple reasons to doubt that this evidence—taken from the human case—is sufficient for establishing a *general* function of consciousness.

To take an analogy, only looking at the function of keratin in reptiles—where it’s used for scales and claws—wouldn’t suffice to establish a *general* function for keratin across taxa, where it’s also used for (e.g.) hooves, fur, tails, and courtship. We therefore aren’t surprised that such a theory would have to be disjunctive. To begin to form one, we’d look wherever we know keratin occurs in nature, and then study the fitness benefits of having keratin in various contexts. In other words, we would need prior knowledge of where keratin occurs to form a disjunctive theory of its function. The same goes for consciousness. Thus, once again, we run into something like the Distribution Catch-22: we require a *function* of consciousness to know its *distribution*, but as long as consciousness has different functions in different taxa, we also must know its *distribution* to infer its functions. To simply assert that the function of consciousness *is* uniform across taxa would therefore be to assert that our standards of evidence in assessing the function of consciousness should be different than they are in assessing the functions of other traits. This *prima facie* incompatible with methodological naturalism, the widely popular view that consciousness is “just another” biological trait, and should be studied like any other.

How could the function-based symmetry proponent respond? With some additional auxiliary assumptions, I imagine she could appeal to Ned Block’s (1995) distinction between *phenomenal* and *access* consciousness. While *phenomenal consciousness* (or P-consciousness) consists in the kind of felt experience we’re concerned about here, *access consciousness* (or A-consciousness) consists instead in sensory information being available to general-purpose processes of inference. Some (e.g. Naccache 2018) have taken P-consciousness to be coextensive with A-consciousness, i.e. all A-conscious organisms are P-conscious, and vice-versa. Now, it is highly plausible that A-consciousness *is* necessary for facilitating the sorts of abilities identified above. For example, some sort of central processing of information is likely required for cross-modal learning, as information from various sense modalities has to be “brought together” for use in decision-making. Therefore, if A- and P-consciousness are coextensive, then though the “function” of consciousness is really a function of A-consciousness, this function will nevertheless *co-occur* with P-consciousness, since A-consciousness exists wherever P-consciousness exists. This is all well and good, though I’ll note that it would be difficult to make the case for the co-extensiveness of A- and P-consciousness while remaining theory-light—i.e. without also arguing for a general theory of consciousness. This could undermine some of the motivation for adopting the function-based approach.

So, in sum: while there are good reasons to endorse a symmetry view if you *assume* a uniform function (or a mostly homogenous family of functions) across taxa, the challenge is to (1) show how we could know such a function exists, and (2) defend that there *is* such a function for consciousness. Alternatively, a case could be made for the co-extensiveness of phenomenal and access consciousness, in which case the “markers” could be the capacities facilitated by access consciousness, which seem to track well with the “markers” identified by many authors. If you endorse the function-based approach and the symmetry view, the choice is yours (Fig. 4b).

A diagram of a question

Description automatically generated with medium confidence

*Fig. 4b: The decision tree for the function-based approach.*

Now, we can put all of these trees together into our final, distilled product.

1. **Conclusion: A (pretty) comprehensive decision tree**

So, can a marker approach exclude? I actually think it can’t, though my thinking so depends on my answers to a number of controversial questions, about which others might disagree. As we’ve seen, there are various “paths” one can take from views about how we identify markers to views about exclusion, and each path comes with “off-ramps” provided by particular answers to these questions. To figure out where you should fall on exclusion in general, you can follow the master decision tree below (Fig. 5). Before you dive in though, keep in mind these three points from the user’s manual.

First, in the introduction, I remarked that my discussion here is not meant to be comprehensive. There may be other approaches one can start from, and those listed may have other “off ramps” provided by auxiliary hypotheses I didn’t consider—that being said, I take the tree to summarize a substantial amount of current disagreement. I hope that future work might extend the tree to accommodate nuances I have missed. Second, it may be that you think that there is more than one legitimate way to identify markers. If so, just complete multiple lines! If you ever get “symmetry,” then you hold the symmetry view. Finally, I have listed papers below most of the ultimate nodes on the tree. This is *not* meant to attribute a view to the paper’s author, but rather to note that the paper itself seems to suggest the associated variety of symmetry or asymmetry. So, for example, though Brian Key’s (2016) paper was, strictly speaking, about *pain*, not consciousness more generally, it plausibly has implications for the latter issue, and was understood to have such implications in its associated commentary. Michael Tye, in addition to holding symmetry-style views on analogy-based approaches, also famously endorses the “PANIC” theory of consciousness, and, despite his paper’s position on this tree, should not be read as rejecting a theory-based approach. Similarly, though Kristin Andrews (2024) supports an asymmetry view of *analogy-based* approaches, Andrews technically holds a symmetry view, as she thinks a mature theory of consciousness *will* be able to exclude—she just doesn’t think this theory will develop anytime soon (Andrews, personal communication).

To end, I’ll run through the tree myself. Personally, I think all three approaches can be used, though due to my own combination of auxiliary hypotheses, I also think they all lead to asymmetry. Theory-based approaches are fine, though since I hold phenomenal consciousness might be multiply realized, I understand *non-disjunctive* theories of consciousness to be theories of consciousness in relevantly physiologically and functionally homogenous groups of organisms; and due to the Distribution Catch-22, I am skeptical of our prospects for a disjunctive theory. Analogy-based approaches are also great, but our initial markers are probably biased, and I don’t think that there’s a theory-neutral way to identify defeaters. Finally, though we may be able to fruitfully investigate the function of consciousness in humans, I don’t think that a uniform “function of consciousness” has been established by our usual evidential standards in evolutionary biology, and I’m not so sure that phenomenal and access consciousness are coextensive.

As concerns about animal minds shift to epistemological matters, we should keep in mind that the question of which animals we can *know* are conscious may require a different answer from the question of which animals we can know are *not* conscious. Future research may benefit from focusing on the issues identified in the tree.

A diagram of a flowchart

Description automatically generated

*Fig. 5: Decision tree linking views about how to identify markers to views about exclusion.*

**Works Cited**

Andrews K, Huss B (2014) Anthropomorphism, anthropectomy, and the null hypothesis. Biology and Philosophy 29:711-729

Andrews K (2018) Mind. In: Lori Gruen (ed) Critical terms in animal studies. Chicago: University of Chicago Press, 234-250

Andrews K (2020) The animal mind: An introduction to the philosophy of animal cognition. Routledge, New York

Andrews K (2022) Does the sentience framework imply all animals are sentient? Animal Sentience. https://doi.org/10.51291/2377-7478.1737

Andrews K (2024) “All animals are conscious”: Shifting the null hypothesis in consciousness science. Mind & Language 39:415-433

Appel M, Elwood R (2009) Motivational trade-offs and potential pain experience in hermit crabs. Applied Animal Behavioral Science. 119:120-124

Baars BJ (2005) Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. Prog Brain Res. 150:45-53

Bayne T, Seth AK, Massimi M, Shepherd J, Cleeremans A, Fleming S, Malach R, Mattingley JB, Menon DK, Owen AM, Peters MAK, Razi A, Mudrik L (2024) Tests for consciousness in humans and beyond. Trends in Cognitive Sciences. 28:454-466

Birch J (2022) The search for invertebrate consciousness. *Noûs.* 56:133-153

Block N (1995) On a confusion about a function of consciousness. *Behavioral and Brain Sciences.* 18:227-287

Brüne M, Brüne-Cohrs U (2005) Theory of mind—evolution, ontogeny, brain mechanisms and psychopathology. Neuroscience & Biobehavioral Reviews. 30:437-455

Crump A, Birch J (2022) Animal consciousness: The interplay of neural and behavioral evidence. Journal of Consciousness Studies. 3:104-128

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

De Weerd C (2024) A credence-based theory-heavy approach to non-human consciousness*.* Synthese. 203:1-26

Dennett D (1995) Animal consciousness: What matters and why. Social Research 62(3): 691-710

Dennett D (1996) Kinds of minds: Towards an understanding of consciousness. Basic Books, New York

Dennet D (2019) Review of Other minds: The octopus, the sea, and the deep origins of consciousness. Biology & Philosophy. https://doi.org/10.1007/s10539-018-9650-2

Elwood R (2021) Potential pain in fish and decapods: Similar experimental approaches and similar results. Front Vet Sci. 8:631151

Feinberg T, Mallatt, J (2016) The ancient origins of consciousness. MIT Press, Cambridge

Godfrey-Smith P (2020) Metazoa: Animal minds and the birth of consciousness. Farrar, Straus & Giroux, New York

Griffin D (2002) Afterward: What is it like? In: The cognitive animal: Empirical and theoretical perspectives on animal cognition. Bekoff M, Allen C, Burghardt G (eds). MIT Press, Cambridge, pp 471-473

Hahn D (1998) Hermit crab shell use patterns: Response to previous shell experience and to water flow. Journal of Experimental Marine Biology and Ecology. 1:35-51

Hume D (2000) A treatise of human nature. Norton DF, Norton MJ (eds), Oxford University Press, Oxford

Jamieson D (1998) Science, knowledge, and animal minds. Proceedings of the Aristotelian Society. 98:79-102

Key B (2016) Why fish do not feel pain. Animal Sentience. 10.51291/2377-7478.1011

Krieger J, Hörnig M, Laidre, M (2020) Shells as ‘extended architecture’: To escape isolation, social hermit crabs choose shells with the right architecture. Anim. Cogn. 23:1177-1187

Leopold D, Logothetis N (1996) Activity changes in early visual cortex reflect monkeys percepts during binocular rivalry. Nature 379:549-553

Lopez D (2008). Buddhism and science: A guide for the perplexed. University of Chicago Press, Chicago

Polger T, Shapiro S (2016) The Multiple Realization Book. Oxford University Press, New York

Putnam H (1967). Psychological predicates. In: Capitan WH, Merill DD (eds) Art, Mind, and Religion. University of Pittsburgh Press, Pittsburg, pp 37-48

Prinz J (2005) A neurofunctional theory of consciousness. Consciousness and Cognition. 9:243-259

Prinz J (2012) The conscious brain: How attention engenders experience. Oxford University Press, Oxford

Prinz J (2018) Attention, working memory, and animal consciousness. In: Andrews K, Beck J (eds) The routledge handbook of philosophy of animal minds, Routledge, New York, pp 185-195

Scheinburg D, Logothetis N (1997) The role of temporal cortical areas in perceptual organization. Proc Nat. Acad Sci USA. 194:3408-13

Schwitzgebel, E (2020) Is there something it’s like to be a garden snail? Philosophical Topics. 48:39-64

Shevlin H (2021). Non-human consciousness and the specificity problem: A modest theoretical proposal. Mind & Language. 36:297-314

Solms M (2022) Truly minimal criteria for animal sentience. Animal Sentience 32(2). https://doi.org/ 10.51291/2377-7478.1711

Tononi G (2004) An information integration theory of consciousness. *BMC Neuroscience*. https://doi.org/101.1186/1471-2202-5-42

Tye M (1997) The problem of simple minds: Is there anything it is like to be a honeybee? Philosophical Studies 88:289-317

Tye M (2016) Tense bees and shellshocked crabs: Are animals conscious? Oxford University Press, New York

Varner G (2012). Personhood, ethics, and animal cognition: Situating animals in Hare’s two-level utilitarianism. Oxford University Press, Oxford

Veit W (2022) Towards a comparative study of animal consciousness. Biological Theory. 17, 292-303

1. Department of Philosophy, Michigan State University, East Lansing, MI, USA [↑](#footnote-ref-1)
2. Program in Ecology, Evolution, and Behavior, Michigan State University, East Lansing, MI, USA [↑](#footnote-ref-2)
3. What I am calling the “asymmetry view” is what de Weerd (2024) has called the “Asymmetry Thesis.” [↑](#footnote-ref-3)
4. There is some difference in how I’m using the term “marker approach” from how previous authors (e.g. Andrews 2024) use the term. These authors would reserve it for analogy- or function-based approaches, though for my purposes, theory-based approaches are “marker approaches” in the relevant sense, because they provide justification for the use of specific markers. The difference here is merely semantic; I hope it won’t confuse. [↑](#footnote-ref-4)
5. Though certain lines of Prinz’s evidence, e.g. pertaining to the importance of intermediate levels of abstraction, come from lesion studies in primates (Leopold and Logothetis 1996; Sheinburg and Logothetis 1997). [↑](#footnote-ref-5)
6. How different do two realizers of a process have to be to fall into different “types”? Here you would consult your favorite theory of multiple realization, e.g., that of Polger & Shapiro (2016). [↑](#footnote-ref-6)
7. Some may object that there *are* ways to know which animals are conscious without a theory (theory-neutral and theory-light approaches [Birch 2022]). We will get to these in due time.

   [↑](#footnote-ref-7)
8. Some may feel that this account gives undue priority to verbal report as evidence for consciousness. If you feel this way, simply replace “verbally reporting creatures” with “creatures whom we take as paradigm cases of consciousness.” [↑](#footnote-ref-8)
9. However, Prinz (2018) does distinguish level-headed mysterianism (LHM, which he favors) from “radical realization mysterianism” (RRM, which he doesn’t). The difference seems to be that LHM organisms all realize the same functional processes, but with different physical realizers, whereas RRM organisms realize different functional processes and yet are nonetheless conscious. However, he also seems to think that no RRM organisms actually exist (I think he must be assuming a pretty liberal theory of multiple realization), so this addendum may just serve as a bulwark against general concerns about panpsychism. Schwitzgebel, who’s uncertain about whether protons are conscious, might endorse RRM. [↑](#footnote-ref-9)
10. Andrews’ own artifact analogy is a COVID test, though I think this isn’t the best metaphor, since testing negative for COVID *does* reduce the probability that you have COVID, in that your posterior probability of illness is lower than your prior probability, given a negative test (Andrews 2024). [↑](#footnote-ref-10)
11. Though this might sound like a global workspace (Baars 2005) or informational integration theory (Tononi 2004), it isn’t. While these theories *identify* consciousness with information-processing of a certain kind, the authors I’m considering merely state that consciousness has this type of processing as its evolutionary function (i.e. they make a claim about the *functional role*, rather than the *nature*, of consciousness). [↑](#footnote-ref-11)
12. See Godfrey-Smith (2020) for a more in-depth discussion of hermit crab shell selection and its meaning for the philosophy of mind. [↑](#footnote-ref-12)
13. Here’s one that might have occurred to you, but which I don’t think is very promising. Organisms often have traits without displaying the typical functions of those traits, for instance due to exaptation (in which a trait with a certain evolved function is appropriated for a different function, e.g. penguin wings for swimming), or atavism (in which a trait remains but no longer fulfills its evolved function, e.g. your appendix). Thus, you might object with the following *reductio*: function-based reasoning would imply that since the typical function of wings is flying, and penguins cannot fly, penguins therefore probably don’t have wings. However, this objection can be overcome for two reasons. First, if *all you knew* about penguins was that they can’t fly, it would be safe to guess they don’t have wings, since most non-flying organisms are non-winged; function-based reasoning doesn’t deliver negative verdicts with *certainty*, but rather delivers them with the usual degree of inductive uncertainty. Second, advocates of this approach might even be able to handle cases of exaptation and atavism more directly, by building in phylogeny clauses, e.g. by claiming that sentience is unlikely in organisms lacking F *and* F-possessing recent ancestors. [↑](#footnote-ref-13)