**HOTT and Heavy:**

**Higher-Order Thought Theory and the Theory-Heavy Approach to Animal Consciousness**

Forthcoming in *Synthese* 203(98).

[Please cite the published version.]

**Abstract**

According to what Birch (2022) calls the theory-heavy approach to investigating nonhuman-animal consciousness, we select one of the well-developed theories of consciousness currently debated within contemporary cognitive science and investigate whether animals exhibit the neural structures or cognitive abilities posited by that theory as sufficient for consciousness. Birch argues, however, that this approach is in general problematic because it faces what he dubs the dilemma of demandingness—roughly, that we cannot use theories that are based on the humancase to assess consciousness in nonhuman animals and *vice versa*. We argue here that, though this dilemma may problematize the application of many current accounts of consciousness to nonhuman animals, it does not challenge the use of standard versions of the higher-order thought theory (“HOTT”) of consciousness, according to which a creature is in a conscious mental state just in case it is aware of being in that state via a suitable higher-order thought (“HOT”). We show this in two ways. First, we argue that, unlike many extant theories of consciousness, HOTT is typically motivated by a commonsense, and more importantly, neutral condition on consciousness that applies to humans and animals alike. Second, we offer new empirical and theoretical reasons to think that many nonhuman animals possess the relevant HOTs necessary for consciousness. Considering these issues not only reveals the explanatory power of HOTT and some of its advantages over rival accounts, but also enables us to further extend and clarify the theory.

**Keywords:** consciousness; nonhuman animals; HOT theory; concepts

1. **Theories of consciousness and animal consciousness**

There has recently been an explosion of interest among philosophers and cognitive scientists more broadly in the question of whether and which nonhuman animals,[[1]](#footnote-1) including invertebrates such as octopods and bees, have consciousness (e.g., Tye, 2016; Carruthers, 2019; Birch, 2022; LeDoux et al., 2023). To be clear, what is at issue here is whether such creatures enjoy *phenomenal* consciousness—the kind of consciousness for which, to adopt Nagel’s (1974) expression, there is something that it is like for one to have it.[[2]](#footnote-2)

A natural way to answer this question would be to pursue a version of what Birch (2022) recently calls the *theory-heavy* approach to animal consciousness. There are, after all, many well-developed theories of consciousness currently debated within contemporary cognitive science, such as global-workspace theory (“GWT”) (e.g., Dehaene, 2014) and integrated-information theory (“IIT”) (e.g., Tononi, 2012).[[3]](#footnote-3) A theory of consciousness seeks to tell us, first and foremost, what consciousness *is*. In doing so, it must account for the difference between conscious and *nonconscious* states—and, in the fullness of time, help us determine the *neural correlates of consciousness* (“NCCs”), or the minimal neuronal mechanisms sufficient for particular conscious states (see, e.g., Koch et al., 2016). On the theory-heavy approach, we start with one of these theories and investigate whether animals exhibit the neural structures or cognitive abilities posited by that theory as sufficient for consciousness.

Birch argues, however, that this theory-heavy approach is in general problematic because it would seem to face what he calls the *dilemma of demandingness*. As he observes, theories of consciousness vary in their degrees of *demandingness—*that is, they differ in terms of the cognitive or neural sophistication required for consciousness. Some theories, such as GWT, would seem to be highly demanding; others, such as midbrain theory (e.g., Merker, 2007), only weakly so. Roughly, the dilemma is that the mechanisms posited by demanding theories may seem reasonable to ascribe to humans, but extend less obviously to nonhuman creatures; by contrast, the mechanisms posited by comparatively undemanding theories may be found in animals, but are not clearly sufficient for consciousness in humans.

While Birch’s dilemma may undermine many or even most extant theories as viable approaches to animal consciousness, our main goal here is to argue that it does not challenge standard versions of the higher-order thought theory (“HOTT”) of consciousness, according to which a creature is in a conscious mental state just in case it is aware of being in that state via a suitable higher-order thought (“HOT”) (e.g., Rosenthal, 1991; 2005; Weisberg, 2011; LeDoux, 2019). Accordingly, we urge that we can and should investigate animal consciousness in the theory-heavy way via HOTT and that this will be a fruitful method.

At first, that HOTT might avoid the dilemma of demandingness may strike many as doubtful, as the theory is often interpreted as being highly demanding (e.g., Passos-Ferreira, 2023; Bayne et al., 2023). After all, it may seem that the ability to have *higher-order* thoughts—that is, thoughts about one’s own mentality—is a *very* sophisticated cognitive achievement. But we argue that HOTT avoids Birch’s dilemma in two ways. First, we argue that, unlike many current theories of consciousness, HOTT is motivated by a commonsense, and more importantly, *neutral* condition that applies to humans and animals alike. We thus propose that HOTs are necessary for consciousness, however cognitively demanding they may be.

But because some theorists may not accept this condition, we argue that HOTT avoids Birch’s dilemma for an additional reason—namely, that HOTs are at the appropriate level of demandingness to allow the theory to sidestep the dilemma. Responding to the dilemma in this way in effect provides a new reply to perhaps the most well-known objection to HOTT, often referred to simply as the *objection from infants and animals* (e.g., Jamieson and Bekoff, 1992; Dretske, 1995; for an overview, see Carruthers & Gennaro, 2020, section 7.2)[[4]](#footnote-4)—roughly, the worry that HOTT rules out the possibility of animal consciousness from the outset.

HOT theorists have replied variously to this objection. Some simply bite the bullet, accepting that only (suitably developed) human beings enjoy consciousness (e.g., Carruthers, 2000; LeDoux, 2019). But this concession strikes many, including us, as too hasty. More plausibly, other HOT theorists have urged that HOTs and the concepts that they employ are not particularly cognitively sophisticated—and so some animals can and do form the relevant HOTs (e.g., Rosenthal, 2005; 2008; Gennaro, 2012; 2018).

We offer here new reasons to think some animals might possess certain types of HOTs. In short, we draw a distinction between types of consciousness—and we argue that there are empirical and theoretical reasons to think that many nonhuman organisms may possess some, but not all, types. Contrary to what many assume, then, HOTT does not rule out animal consciousness. Indeed, we argue that it predicts and explains the distribution of types of consciousness in the animal kingdom in a more straightforward manner than rival theories.

To be clear, while we believe that HOTT is well supported both theoretically and experimentally (see, e.g., Rosenthal, 2005; Lau & Rosenthal, 2011; LeDoux, 2019), our goal here is not to argue for HOTT *per se*.[[5]](#footnote-5) But our intent is not simply to respond to the dilemma of demandingness either. Rather, we urge that considering how HOTT avoids this dilemma not only reveals the explanatory power of HOTT and some of its advantages over other accounts of consciousness, but also enables us to further extend and clarify the theory.

We begin in Section 2 by motivating the theory-heavy approach to animal consciousness—as well as the dilemma of demandingness that some of its applications face. Then, in Section 3, we argue that HOTT avoids the dilemma because of its organism-neutral commonsense grounding. In Section 4, we argue that, even if one were not persuaded by this motivation, HOTT avoids the dilemma because at least *some* HOTs are not overly cognitively demanding. We close in Section 5 with some brief reflections on the search for animal consciousness.

1. **The theory-heavy approach and the dilemma of demandingness**

Why use a *theory* to investigate nonhuman consciousness in the first place? After all, one might think that we might instead opt for what Birch (2022) calls a *theory-neutral approach*, on which we do not use any particular account of consciousness to determine which creatures have it; instead, we simply, as Birch puts it, “build up a list of the behavioural, functional and anatomical similarities between humans and non-human animals, and use arguments from analogy and inferences to the best explanation to settle disputes about consciousness” (2022, p. 134). But Birch himself persuasively argues that such an approach fails. We cannot know if our analogies are apt—for example, whether the absence of certain behaviors or neural structures are defeaters for an analogical inference—without at least some basic theoretical grip on what consciousness involves.

Birch nonetheless doubts that we require a full-blown theory of consciousness to tell us what evidence would reveal animal consciousness. Rather, he urges that we pursue a version of what he calls a *theory-light approach*, which seeks to identify *minimal* credible markers of consciousness, purportedly compatible with a range of theories. Birch and colleagues (2020) propose, for example, that such a marker might be evidence of so-called *unlimited associative learning* (“UAL”), which is “a form of associative learning that allows a system to learn about itself and the world in an open-ended, exploratory manner” (Halina et al., 2022, p. 63).

As others have argued (e.g., Brown, LeDoux, & Rosenthal, 2021; Halina et al., 2022), however, it is unclear what the minimal credible markers of consciousness might be—and indeed whether there are such credible markers. This is because there is growing experimental evidence that many, and perhaps even most, behaviors can be driven by *nonconscious* mental states or processes (for overviews, see, e.g., Lau & Rosenthal, 2011; LeDoux, 2019). In the neurological condition known as blindsight, for example, individuals with damage to visual cortex claim not to be able to see stimuli presented to parts of their visual field, but are nonetheless above chance at discriminating them, which indicates that they see, but do not consciously see, those stimuli (e.g., Weiskrantz, 1986). And blindsight-like behavior has been experimentally demonstrated in nonhuman animals such as macaque monkeys (e.g., Cowey & Stoerig, 1995; Yoshida & Issa, 2015). So, the mere fact that an animal engages in certain behaviors is not sufficient to demonstrate that the creature is in conscious states. As LeDoux recently puts it, our theoretical “default position” therefore ought to be that “behaviour is controlled nonconsciously until proven otherwise” (2019, p. 328).[[6]](#footnote-6)

Indeed, as regards Birch’s own proposed theory-light marker, Brown, LeDoux, and Rosenthal (2021) persuasively argue that UAL, or at least some of the component behaviors of it, may be driven by nonconscious states too. For such reasons, we think a theory-heavy approach, at least in the near term, is justified.

Birch (2022) nonetheless provides what may seem to be a decisive reason to think that the theory-heavy approach is problematic—namely, the dilemma of demandingness. Here is how he summarizes it:

The general dilemma is this: strong sufficient conditions for consciousness will not get us very far in making inferences about cases other than humans who can report their experiences, if they get us anywhere. Yet, as we formulate increasingly weaker conditions, the evidence from humans that they amount to a sufficient condition becomes weaker, and the positive case for animal consciousness becomes correspondingly weaker (2022, p. 138).

In other words, we can formulate the dilemma as involving the following two horns:

***The necessity horn***: the more demanding a theory’s sufficient conditions for consciousness in humans, the less likely it is that evidence that those conditions are met is necessary for attributing consciousness in animals.

***The sufficiency horn***: the less demanding a theory’s sufficient conditions for consciousness in humans, the less likely it is that evidence that those conditions are met is sufficient for attributing consciousness in animals.

To see the dilemma in action, consider Birch’s two central examples of theories of consciousness: GWT and midbrain theory. Roughly, these theories hold:

**GWT**: An animal A is in conscious mental state M just in case M is “in” the global workspace, a frontoparietal network in the cortex that possesses long-range neural connections to, and thus makes information available (via “global broadcast”) to, a variety of neural subsystems responsible for a range of psychological functions, such as action planning and verbal report.

**Midbrain theory**: An animal A is in conscious mental state M just in case M is suitably neurally realized by the midbrain.

GWT is typically understood to be a highly demanding theory insofar as it requires that organisms have rather evolutionarily recent cortical structures responsible for czomplex cognitive processes and behaviors. As a result, the case for the sufficiency of a state’s being in the global workspace for consciousness may seem rather clear.

But demanding theories thereby face the necessity horn of the dilemma. In the case of GWT, bees do not have global workspaces, much less cortex. But, Birch observes, appealing to GWT to arrive at the conclusion that bees do not have consciousness seems too quick. There will always be a question whether whatever mechanism a demanding theory claims is sufficient for consciousness in healthy adult human brains, such as a state’s presence in the global workspace, is also *necessary* for consciousness in nonhuman animals. It seems at least open that bees’ complex nonverbal behavior (such as their well-known communicative waggle dances) supports the view that bees have conscious states that occur outside of a global workspace.

Midbrain theory is, by contrast, a fairly undemanding theory. The midbrain is an evolutionarily old and subcortical structure; many organisms would seem to have neural mechanisms comparable to human midbrains. It is therefore open on the theory to attribute consciousness to creatures fairly low on the phylogenetic ladder (including, for example, insects; see, e.g., Barron and Klein, 2016).

But such undemanding theories thereby face the sufficiency horn: that it is not at all clear that evidence of such undemanding conditions, such as midbrain structures, is enough to attribute consciousness to such creatures, as opposed to complex behavior driven by nonconscious processes. After all, when cortex is heavily damaged in humans, one cannot elicit most behaviors that would reliably determine whether the individual is in conscious mental states—so it is hard to be confident that preserved subcortical midbrain structures, or whatever structures undemanding theories posit, really are sufficient for attributions of consciousness in humans, and thus in animals as well.

What about HOTT? A bit more carefully, we take standard versions of the theory to hold:

**HOTT**: An animal A is in a conscious mental state M just in case A is aware of itself as being in M via a suitable HOT M\*.

Moreover, on standard versions of HOTT, HOTs are theorized to be ordinary assertoric thoughts with conceptual contents of the form: <I am in mental state M>. So, such thoughts are not only conceptual, but deploy both concepts of *mental states* and of *oneself* as their subject.

Perhaps unsurprisingly, HOTs are thus often interpreted as being highly cognitively demanding. Indeed, HOTT is often interpreted as being at least as demanding as GWT, if not more so (see, e.g., Bayne et al. 2023, p. 4). In turn, it might seem obvious that the theory faces the necessity horn of the dilemma of demandingness. We argue now, however, that it does not.

1. **HOTT, organism-neutrality, and the necessity horn**

While HOTT is supported by a range of considerations (see, e.g., Rosenthal, 2005; Lau & Rosenthal, 2011), arguably its main motivation is the commonsense observation that, if one is in a mental state, but in no way aware of being in it—as in a case of a subliminal perceptual state in blindsight—then that state is nonconscious. Most people would agree with this. But that statement is logically equivalent to the claim, which is thus implicit in folk psychology, that Rosenthal (e.g., 2005, p. 4) has called the *transitivity principle* (“TP”):

**TP:** An animal A is in a conscious mental state M only if A is somehow aware of being in M.

The TP specifies a necessary condition on consciousness. So-called higher-order (“HO”) theories in general typically seek to understand how the TP is *implemented*, insofar as they hypothesize what the mechanism of consciousness might be—that is, the mental states in virtue of which one is aware of one’s mind that are not just necessary, but also sufficient for consciousness.

HO theorists disagree about the nature of these HO states and mechanisms. While HOTT proposes that they are thoughts of a certain kind, other HO theories have over the years posited other types of HO state, such as quasi-perceptual states (e.g., Lycan, 1996) or mental “discriminator” or “pointer” states (e.g., Lau, 2022).[[7]](#footnote-7) Whatever the nature of the relevant HO states, such states are HO insofar as they are about or represent other mental states. First-order (“FO”) states, by contrast, are states that concern not one’s mind, but features of the world, including one’s body.

Details about HO theories aside, what we highlight now is that the TP specifies an *organism-neutral necessary condition for consciousness grounded in common sense*. ‘Organism-neutral’ here means that the TP applies not simply to human consciousness, but also to *any form* of consciousness. Consequently, the TP may guide inferences about consciousness in any creature, and this is precisely what is needed to confront Birch’s dilemma. Furthermore, since the condition is grounded in common sense, it is also *theory* neutral. The TP is supposed to serve as the starting point for *any* theory of consciousness.

That HOTT is motivated by an organism neutral condition sets it and other HO theories apart from many other prominent theories of consciousness in the literature. Of course, other theories such as GWT *aspire* to organism neutrality, insofar as they are put forward as theories of consciousness *tout court*. But many theories fail to have good grounds to apply to nonhuman organisms because the main strategy by which they are developed is the *contrastive* or comparative methodology (see, e.g., Baars, 1988). Roughly, experimenters stipulate some operationalized behavioral markers of consciousness in humans—such as verbal report—and then compare the behavioral or neural differences between human participants in what they deem to be conscious and nonconscious trials on the basis of these selected markers (for a review, see, e.g., Michel & Morales, 2019, p. 496). But theories that depend on this contrastive methodology thereby do not begin with an *independent* organism-neutral necessary condition for consciousnesssuch as the TP. Such theories may thus aim to apply to any type of creature, but actually only tell us what might be necessary and sufficient for consciousness in the studied case—usually humans. They do not furnish predictions for what conditions might be necessary, much less sufficient, for consciousness in nonhuman animals. And this is precisely why they face the necessity horn of Birch’s dilemma.[[8]](#footnote-8)

We acknowledge, however, that it is often difficult to determine what is really part of common sense, and furthermore that not all theorists accept the TP (e.g., Dretske, 1995; Brown, 2015). Theorists who reject the TP often claim that we can instead pick out conscious states in a simple and commonsense way using Nagel’s terminology: as states for which there is something that it is like to be in them. But, as Berger and Brown (2021, p. 639) recently urge, it is arguable that the standard Nagelian way to characterize consciousness coincides with the TP. After all, if one is in a mental state, but in no way aware of being in it, then there is nothing that it is like for the organism to be in that state. But that is simply equivalent to the claim that there is something that it is like for an organism to be in a state only if the organism is somehow aware of being in it. So, it is open to understand conscious states not simply as states for which there is something that it is like to be in them, but also as states that we are suitably aware of being in. The TP is thus compatible with using ‘what it’s likeness’ as a way to identify conscious states.

One might nonetheless object that the TP does not address the necessity horn because it may seem to depend on considering the human case—because human beings assent to it. Why, then, think that such inner awareness is required for nonhuman consciousness?[[9]](#footnote-9)

Here, it’s important to emphasize that the TP is not meant to be restricted to human experience. Rather, it is meant to capture our commonsense grip on consciousness *in general.* It starts not from reflection on our human impression of consciousness from the first-person perspective, but rather from reflection on the nature of nonconscious states from the *third-person* perspective. Recall, the claim is itself typically introduced not as the TP, but as its contrapositive—that, if a creature of any type is in a mental state, but in no way aware of being in it, then that state is plainly nonconscious. But this claim captures our general psychological understanding of *non*consciousness, which in turn delivers an organism-neutral understanding of consciousness: the TP.

Insofar as it does not depend on the first-person human perspective, the TP has an advantage over the Nagelian description of conscious states as states for which there is something that it’s like to be in them, which does so depend on first-person impressions. But note that the ‘what it’s like’ description is simply assumed to be organism-neutral too—indeed, Nagel (1974) himself famously assumes it describes both human and bat consciousness. So, if the Nagelian description of consciousness is organism neutral, so is the TP. Just as it is impossible for a creature to be in a conscious state but there be nothing that it is like for the creature to be in it, so too it is impossible for a creature to be in a conscious state but the creature not be suitably aware of being in that state.

In light of the foregoing, perhaps one will grant that the TP is organism neutral, but reply that we have not yet defended HOTT against the necessity horn of Birch’s dilemma, since the TP concerns the *necessary* conditions for consciousness, and it is the theory-based *sufficient* conditions for consciousness that the necessity horn targets. That is, one might agree that *some* form of HO awareness is necessary for consciousness, but question whether some *specific* form of cognitively demanding HO awareness that is theorized to be sufficient for consciousness in humans, such as that which is implemented by HOTs, is really necessary for consciousness in animals.

But HOTT still avoids the necessity horn because the necessity of HOTs for consciousness does not depend on consideration of the TP alone. It is true that the TP establishes only the necessity of some form of HO awareness, but *additional theorizing* and argument establishes the necessity of that HO awareness’ being implemented by HOTs, by establishing the sufficiency of such awareness for consciousness. There are many ways that this additional reasoning might go. Rosenthal arrives at HOTT in effect by eliminating other possible options for the nature of the suitable HO state—what we called M\* above. For example, he (e.g., 2005, pp. 182-3) argues against the quasi-perceptual states of Lycan’s (1996) higher-order perception (“HOP”) theory because, among other things, there is no HO sensory modality within which to locate such HOPs. Of course, Rosenthal does not, nor can we, consider all possible alternative analyses of suitable HO states. Considering such issues would take us too far afield and is in any case unnecessary. Our point is that these further considerations may also proceed in an organism-neutral way. That HOPs would require a HO sensory modality is a point about perception generally, not a point about human perception. The TP, plus some additional argumentation, may thus deliver the result that HOTs are necessary in *any type* of creature.

If one is persuaded by the forgoing remarks about the TP, then our case is closed. Still, we recognize that there may be additional reasons that one might question or even deny the TP—for example, one might follow Block (2007) in arguing that there is experimental evidence of phenomenological “overflow,” or cases of conscious states to which subjects lack any form of cognitive access. While we believe such considerations are far from decisive (on the case against overflow, see, e.g., Brown, 2012), our goal in this section has been less ambitious than that of *establishing* HOTT. Instead, one can understand our arguments here to be *conditional*—namely, that *if* one accepts the TP, then HO theories, and in particular HOTT, do not face the necessity horn of the dilemma of demandingness.

But even if one rejects the TP, we urge that HOTT still avoids the necessity horn because, contrary to what many assume, HOTs are *not* so cognitively demanding so as to rule out that many nonhuman animals might possess them. It is to that issue that we now turn.

1. **The cognitive demandingness of HOTs**

The idea that animals likely do not or cannot possess HOTs is often driven by relatively stringent assumptions about the nature of thoughts and their conceptual constituents. As an illustration, one common assumption, going back at least as far as Descartes (1988), is that genuine thought or concept possession requires language; since nonhuman animals lack linguistic abilities, they thereby lack concepts (see also, e.g., McDowell, 1994). And it this reason—that animals do not possess language—that motivates some HOT theorists to deny or at least question that they have consciousness (e.g., LeDoux, 2019).

Independent of HOTT, however, there are many reasons why denying nonhuman animals can be and often are in conscious states is questionable. For one thing, common sense does seem to hold that many animals have conscious perceptions and conscious sensations, such as visual experiences of colors or conscious pains. Likewise, that blindsight-like behavior can be induced in some creatures, such as macaque monkeys, suggests that their ordinary vision is conscious.

Fortunately for HOTT, the growing consensus in cognitive ethology is that language is not necessary for concept possession and that at least some nonlinguistic animals do enjoy conceptual capacities (e.g., Newen & Bartels, 2007; Wild, 2021). There is a wealth of growing evidence that animals engage in many types of complex behaviors that would seem to require conceptualization. Take just one example: the striking evidence that scrub jays can not only remember where they have stored food after several days, but also recover stored food in a way that is sensitive to a trade-off between how preferable that food is, and how likely it is to have decayed in the time interval between storage and retrieval. If the time interval is short, jays will go for the preferable but perishable food (e.g., wax moth larvae); whereas if it is long, they will retrieve the less preferable, yet non-perishable food (e.g., peanuts) (e.g., Clayton & Dickinson, 1998). This suggests that jays recognize the stored food items as belonging to distinct categories with differing timelines for decay, and use this information to plan and direct their behavior even in the absence of the relevant food targets. And this indicates conceptual thought, albeit perhaps of a comparatively rudimentary form.

Of course, the nature of concepts, concept possession, or concept acquisition are complex topics; in what follows, we won’t defend any particular account of each (for a review of various theories, see the essays in, e.g., Margolis & Laurence 2015). But we are sympathetic with accounts of the nature of concepts and concept acquisition that other HOT theorists, such as Gennaro (2012, especially Chapters 6 and 7), have developed.[[10]](#footnote-10) For example, Gennaro defends what he calls ‘CONPOSS’ (for concept possession), on which:

(CONPOSS) Whenever a subject S has an empirical concept C that is applied to some object (or property or relation) in experience e, S must at minimum (a) be able (to some extent) to discriminate instances of C’s from non-C’s, (b) be able (to some extent) to recognize or identify instances of C by virtue of at least some central feature of the objects or properties in e, and (c) be able to include the concept C in at least some intentional states that S has (2012, p. 144).

Likewise, we understand concept possession as minimally requiring the ability to deploy repeatable mental representations that underwrite certain discriminative and recognitional abilities and that can be deployed in thought. On this minimal conception of concepts, and according to many theories of concepts, some animals, such as scrub jays, do possess them.

But the above examples of animal thoughts are FO insofar as they are of or about elements of their ambient environment. We thus proceed on the assumption that what is at issue is not whether animals possess *any* concepts or thoughts, which many would be willing to grant, but whether they may have *HO* concepts and thus HOTs. That strikes many as a bridge too far (e.g., Jamieson and Bekoff, 1992; Dretske, 1995; for an overview, see Carruthers & Gennaro, 2020, section 7.2), but we now argue that it is not.

* 1. **Concept possession and types of consciousness**

At first, one might deny that animals possess HOTs, not because they lack concepts altogether, but because they lack the capacity for any type of self-awareness—and so the capacity to apply mental-state concepts to themselves. Whether animals have self-awareness, or have that self-awareness via applying a concept of the self, is plainly another huge debate that we cannot settle here (for review, see, e.g., DeGrazia 2019). But we briefly note that there is also growing ethological evidence that many nonhuman animals do exhibit some degree of self-awareness. There are many experimental means of assessing this, perhaps the most well-known of which are varieties of the so-called *mirror self-recognition test* wherein creatures identify markings on themselves as they are presented in mirrors, suggesting that they can recognize themselves *as themselves* as opposed to others. And many creatures, including apes, elephants, and even creatures as low down on the phylogenetic ladder as cleaner fish, have been demonstrated to have this sort of awareness (on elephants and fish, see respectively, e.g., Plotnik, 2006; Kohda et al., 2023). So, we think it is provisionally reasonable to proceed on the assumption that animals not only possess concepts, but also could possess self-concepts.

Even if creatures can conceptualize themselves as distinct from others, however, such concepts may still seem to be FO insofar as perhaps they involve only conceptualizing aspects of their bodies rather than their mental states. Moreover, there may seem to be equally good experimental reasons to deny animals have HOTs.

Perhaps the main reason is the well-known experimental evidence that many nonhuman animals, even sophisticated primates such as chimps, bonobos, and orangutans, fail standard tests of theory of mind (“ToM”), such as nonverbal false-belief tasks (see, e.g., Call & Tomasello, 1999). There are many varieties of such tests. A classic version involves primate participants watching a human trainer (the “hider”) hiding a treat in a box while another trainer watches (the “communicator”). The communicator then leaves the room, during which time the hider switches the two boxes. Upon returning, the communicator places the token on the box in which they falsely believe the treat is—the one they observed the hider placing it in. When given the choice, primates persist in choosing the box that the communicator has placed the token on, suggesting that they are not able to register that the communicator falsely believes there to be a treat in it. Variations of this task have produced consistent results (see, e.g., Krachun et al., 2009). That most nonhuman animals do not pass such tests has suggested to some, including some HOT theorists (e.g., Carruthers, 2000), that (most) animals lack the mental-state concepts required for HOTs.

Other HOT theorists have replied to this concern. Gennaro (2012, p. 249ff), for example, has urged that, even if nonhuman animals cannot attribute mental states to *others*, they may still be able to attribute them to *themselves*, which is all that is required for the relevant HOTs. But such a proposal naturally requires an explanation for this asymmetry. Gennaro maintains that it is reasonable to think that creatures would have a better understanding of a mental-state concept when applying it to oneself than when applying it to another. But this is questionable. Presumably, if a given organism can attribute to themselves a mental state, then they understand at least some of the behaviors that the state is connected with, since they too would exhibit those behaviors when in it. And if so, then it is unclear why they would not be able to attribute that state to others just as easily as they self-attribute it. We thus agree with Gennaro that nonhuman animals possess HOTs, but we now offer an explanation of why nonhuman animals cannot pass standard ToM tests that does not evoke any asymmetry between their capacity to attribute mentality to themselves and others.

Crucially, it has not yet been adequately observed that, while ToM tasks all test for the ability to attribute mental states to others, they often differ with respect to the different *types* of mental state they test for the ability to attribute. Though current evidence seems to show that nonhuman animals regularly fail false-*belief* tasks, there is growing confirmation that animals can and do successfully attribute other states, such as *perception*, to one another (e.g., Hare et al., 2000; Call & Tomasello, 2008; Whiten, 2013). As Gennaro helpfully summarizes some of this work:

Rhesus monkeys preferentially attempt to obtain food silently only in those conditions where silence was relevant to obtaining the food undetected. While a human competitor was looking away, monkeys would take grapes from a silent container, thus apparently understanding that their human competitors would hear the noise otherwise. Subjects reliably picked the container that did not alert the experimenter that a grape was being removed. This suggests that monkeys take into account how auditory information can change the cognitive state of the experimenter (2018, p. 11).

That is, the current experimental evidence suggests that, while nonhuman animals may not be able to attribute beliefs, and hence lack the concept BELIEF, they may possess concepts of HEARING or SEEING and can thereby attribute such states.

Notice, however, that if HOTT is true, it may be that creatures vary in terms of which *types* of states they can attribute to themselves and thereby which types of conscious states they can be in. According to HOTT, the types of conscious states that creatures can be in depends on the types of concepts that they possess and that figure in their HOTs. A FO thought, for example, would be conscious in virtue of one’s being in a suitable HOT with a content that deploys the concept of THINKING—a content such as <I think that that there is food present>. By contrast, a FO perceptual state would be conscious in virtue of one’s being in a suitable HOT that deploys the concept of SEEING, such as a HOT with the content <I see red>. So, on HOTT, it may be that nonhuman animals are unable to form HOTs about beliefs, and thus do not have conscious beliefs, though they are able to form HOTs about perceptual states, and thus may have perceptual experiences.

Indeed, independent of HOTT, there is reason to think there might be a difference in the distribution of types of consciousness in the animal kingdom. As noted, common sense and some experimental evidence indicates that animals are often in states of *perceptual* or *sensory* consciousness. But the case is far less clear with regard to other types of state. Folk psychology seems silent regarding whether animals ever have, for example, conscious thoughts, doubts, or intentions. And there does not seem to be any corresponding experimental evidence regarding such states. To our knowledge, there is no evidence that we can somehow experimentally render the conscious thoughts of macaque monkeys into some form of subliminal or implicit cognition.

Any theory of consciousness thus requires an explanation not only of why animals possess consciousness in general, but whythey may have the capacity to be in some but perhaps not all types of conscious state. So, if HOTT is on the table, the question isn’t: can animals have HOTs—and thus consciousness? Rather, the questions are: which types of HOTs can animals have, and thus which *types of conscious state* can animals be in?

In that way, HOTT arguably gives us more to work with when exploring animal consciousness than many other theories. To see why, consider standard versions of GWT. One problem for GWT, which contributes to its facing the necessity horn, is that we have little independent theoretical handle on what a global workspace is outside the human case—and thus not only whether they are necessary for consciousness, but also what it would even be for a nonhuman animal to have one. Consequently, Carruthers (2019), a proponent of GWT, concludes that the question of whether nonhuman animals have conscious is actually indeterminate! But cognitive science arguably has a far better grip on the nature of concepts in general, and the nature of concepts can arguably be settled independently of consciousness. After all, if there is unconscious thought, as many believe, then concepts can occur without being conscious. Exploring which concepts, and thus which HOTs, animals have would thus seem to be a tractable question—one that need not lead us into indeterminacy about animal consciousness.

Moreover, if indeed it is the case that there is a difference in the distribution of the types of conscious states creatures across the animal kingdom can enter into, HOTT is in a strong position to explain that distribution in a way that other theories are not. On some versions of GWT, for instance, sensory states can enter the global workspace, but states such as thoughts or beliefs cannot—so these states are never conscious. On such views, the global neuronal workspace is a sensory-based system (see, e.g., Carruthers, 2019). Such views simply deny the possibility that *any* creatures, including humans, can have conscious thoughts or beliefs. HOTT, by contrast, not only leaves this possibility open, but also offers a principled reason for doing so that does not rest on details about the functional or neuroanatomical properties of the *human* brain. Alternatively, for those versions of GWT that accept conscious thought, we require an explanation of why many nonhuman animals’ perceptions but not thoughts ever enter the global workspace. But it is not clear what such an explanation might look like.

To avoid the necessity horn of the dilemma, however, HOTT must not only explain how animals could be in different types of conscious state, but also *why* it is reasonable to think that they may have HOTs about the types of state that they do. We propose such an account in the next section.

* 1. **Acquiring different types of HOT**

As noted, other HOT theorists have previously offered reasons to think that nonhuman animals might possess HOTs. Gennaro (2012, especially Chapters 7 and 8), for example, has persuasively argued that it is reasonable to think nonhuman animals might have HOTs in virtue of their possession of *simpler* concepts of mental states—that is, concepts of which they have a less complete grasp relative to humans. After all, even restricting ourselves to the human case, one’s mastery or grasp of a concept may admit of degrees, such that one person’s grasp of a concept over time, or relative to another’s grasp of the same concept, might be more or less complete. To use an example of Gennaro’s, a child’s grasp of the concept TREE is at best partial compared to a typical adult’s grasp of the concept, and a typical adult’s grasp is less complete than a botanist’s grasp.

Gennaro does not give a full account of what it is for one’s grasp of a concept to be more or less complete. And, again, we will not be assuming any particular theory here either. But Gennaro (2012, p. 248) does suggest that one’s degree of grasping a concept minimally relates to the connections that one is able to draw among distinct concepts. On this view, we sometimes might not “see” the connection between two concepts, or thoughts containing those concepts, because we have only a partial grasp of the concepts involved. Along these lines, for instance, while the child is able to draw connections among their concept of TREE and the concepts of TRUNK, BRANCHES, and LEAVES, they might not be in a position to draw connections among TREE and the concepts of ROOTS and HABITAT as an adult might, nor the concepts of CONIFEROUS and DECIDUOUS as a botanist presumably would.

We agree with Gennaro that one’s grasp of a concept can be less complete than another’s—as a nonexpert’s grasp of TREE is less complete than a botanist’s. But Gennaro’s view would not on its own explain why certain *types* of consciousness would be available only to some types of creatures.

Gennaro does, however, provide another oft-noted insight about concept acquisition that is relevant—namely, that creatures seem to acquire different types of concepts *in stages* (e.g., 2012, p. 199).For example, he notes that a comparatively simple concept, such as the concept of a table, is in most cases acquired far earlier in development than complex scientific concepts such as QUARK.

We propose that a natural explanation of this fact is that some concepts are simply *easier* to grasp in more complete ways than others, due to their relative simplicity. And we note that these two dimensions, ease of acquisition and degree of mastery, are arguably independent of one another. It’s doubtless easier for a child to acquire the concept TABLE than it is for the child to acquire QUARK, though the child’s eventual mastery of both concepts may be only comparatively partial compared to an adult’s or an expert’s.

We naturally cannot offer a full account here of the relative ease of acquiring certain concepts. But how easily a concept can be acquired at least in part depends on how many connections between the target concept and other concepts one must observe to acquire it, as well as how rich or complex those conceptual connection are.[[11]](#footnote-11) A child, for example, may know and be able to say the word ‘quark’, but does not come to possess a serviceable concept of QUARK until the child possesses sufficiently many other concepts, such as ATOM or CHARGE, and understands their connection in some albeit incomplete manner. In that way, QUARK is a comparatively difficult concept to develop even a rudimentary grasp of, as it requires noting connections between many other sophisticated concepts that themselves require noting many conceptual connections. The concept TABLE, by contrast, is comparatively easy to acquire, if not master, as one may only need to appreciate certain simpler connections, such as to the concepts SURFACE and FURNITURE.

Returning to HOTT, HOT theorists have observed that there are similarly good reasons to think that the concepts that figure in some HOTs are somehow easier to acquire than the concepts that figure in others (e.g., Rosenthal, 2005, p. 40; Gennaro 2012, p. 226). That is, they maintain that concepts of perceptual states and bodily sensations are less complex, and hence easier to acquire, than concepts of cognitive states such as thoughts.

But we can say a bit more about why acquiring concepts of states such as belief requires a more complex system of concepts than that required for concepts of sensory states. Whatever else one thinks about sensory or perceptual states, such states are constrained by sensory processing—and hence provide access to restricted families of properties that can be characterized by so-called *quality spaces* (e.g., Quine, 1960; Rosenthal, 1991; 2005). Vision, for example, provides access to perceptible colors, which can be mapped onto the quality space of colors—also sometimes known as the color solid—which is a three-dimensional space characterized by the dimensions of hue, saturation, and brightness.

To acquire basic grasps of concepts of perceiving, then, one must appreciate only that creatures have a way of tracking such restricted ranges of perceptible features in their environments by their sense organs. That is, one must appreciate connections between a small set of other easy-to-master concepts such as CREATURE, EYE, and COLOR, where the latter concept only involves connections to a limited range of other concepts, such as RED and GREEN. Though no one, including humans, possesses concepts of each of the members of these quality-space families, possessing even a very partial understanding of the dimensions of the relevant spaces enables creatures to form many fairly rich concepts of the members of the family. Once one possesses SEE and RED, it’s only a small step to SEE DARK RED or SEE ORANGE TO THE RIGHT OF RED. These are precisely the kind of concepts that we maintain many nonhuman creatures might grasp, however crudely.

Similar remarks can be made about perceptual states in other modalities, such as states of audition or olfaction, and indeed arguably about sensory states such as pains, tickles, and itches. Young and colleagues (2014), for example, argue that smells vary in ways that arguably can be captured by a multidimensional quality space; similarly, Rosenthal (2005, p. 140ff) argues that pains vary among one another in terms of being sharp, dull, burning, and so forth. These simple relationships between perceptible or sensible qualities are precisely the kind of thing that we are urging a creature could conceptualize comparatively easily. Indeed, this is even more plausible if one grants that animals may in general have more rudimentary grasps of any such concepts than humans.

By contrast, the concepts that figure in HOTs about other types of cognitive state, such as beliefs, are vastly more complex and thus far more difficult to grasp. While perceptual states are constrained by sensory processing and so target restricted families of properties, belief states are not so constrained—and so can be about virtually anything. One can think not only that there is red before one, but also that food is present. And beliefs stand in complex inferential relations to one another. Beliefs interact in complicated ways with desire to drive action, and beliefs lead to other beliefs via reasoning. Acquiring a concept of belief, or desire, thereby requires appreciating (at least to some minimal degree) such states’ comparatively far more complex roles. One cannot have the concept of a belief about food without the concept FOOD, much less without the concept BELIEF, which in turn arguably requires concepts such as TRUE or FALSE, and the ability to understand at least in part how beliefs interact with desires, perceptions, and other mental states.[[12]](#footnote-12)

Of course, there is much more to states of seeing or believing than this. States of seeing cause beliefs, for example, and they have various neural bases. But one need not appreciate *those* connections to minimally possess the concept SEEING. By contrast, to even minimally grasp BELIEF, one must appreciate much more.

This is all to say that there are good reasons to think some nonhuman animals plausibly can possess rudimentary concepts of certain types of states, such as perceptual or sensory states, and thus may possess those types of consciousness, but not even rudimentary concepts of other types of states, and thus may not possess those other forms of consciousness. In other words, HOTT predicts and explains a difference in distribution of types of consciousness in the animal kingdom, which we have some initial evidence for. It is not clear that other theories of consciousness make such predictions or provide such an explanation.

In sum, we’ve argued that, contrary to appearances, HOTT is not all that demanding a theory. Does that mean that it faces not the necessity, but the sufficiency, horn of Birch’s dilemma? No. We acknowledge that other variants of HO theory might. One might think, for example, that Lau’s (2022) version of HO theory falls on the low end of the demandingness spectrum, since it holds that the relevant HO mechanism is a subpersonal, nonconceptual “discriminator” that targets, but does not represent, a FO state. But whether or not Lau’s view faces this worry, the comparatively more cognitively demanding HOTT is safe. Even if HOTs are not so demanding that they rule out animal consciousness, we take them to be sufficiently demanding that they do not face the sufficiency horn in the way that other *very* undemanding theories, such as midbrain theory or IIT, arguably do.

1. **Conclusion**

Whatever the merits of HOTT in the human case, the theory is often thought to provide inadequate accounts of nonhuman consciousness. Here, we have argued that this is not the case—that HOTT can and should be used in what Birch calls the theory-heavy way to investigate whether and which creatures have consciousness.

In particular, we have argued that HOTT avoids the dilemma of demandingness because, unlike many other extant theories of consciousness, it is grounded in an independent organism-neutral motivation. So, however demanding HOTs might be, there is reason to think that they are necessary for nonhuman consciousness. But even if one rejects this organism-neutral motivation, we have argued that HOTT avoids the dilemma because there are good reasons to think that nonhuman animals possess certain types of HOTs, and thus may enjoy certain types of consciousness, even if they do not possess other types of HOTs, and thus may not experience the full range of types of consciousness that humans do.

One might worry: even if HOTT is not that demanding, so that many types of creatures—from elephants to fish—have HOTs, the theory remains sufficiently demanding that many creatures, such as insects, cannot have them. It may thus seem that HOTT implausibly denies that insects enjoy consciousness, doing violence to the intuitions of many that such simple organisms can be in conscious states too.

We unfortunately cannot do full justice to this objection here. But we’ll make one observation, which we hope may address it to some extent. According to most theories of consciousness, we must distinguish conscious states, such as perceptual experiences, from nonconscious and *subpersonal* states, such as retinal states that occur in the eye. But most theories also distinguish these states from their nonconscious, but *personal-level* counterparts, such as subliminal visual states that occur in primary visual cortex. That is, on many theories of consciousness, one can not only consciously see, but also nonconsciously see—and not simply nonconsciously *register* information about the environment. Although the personal/subpersonal-level distinction is a vexed one (see, e.g., Drayson, 2012), personal-level perceptual states, whether conscious or not, would seem to have much in common. They minimally modally encode information about one’s environment in a way that makes that information available for action-guidance, whereas subpersonal retinal states do not (see, e.g., Berger & Mylopoulos 2021; Shepherd & Mylopoulos 2021). It is thus reasonable to operate with a tripartite distinction among conscious mental states, nonconscious personal-level mentality, and nonconscious subpersonal states.[[13]](#footnote-13)

Because many in the history of philosophy such as Descartes (1988) have assumed that all mentality is conscious, the question that has occupied many theorists is whether or not nonhuman animals such as insects possess *consciousness*. But once we distinguish the conscious and the mental, we can see that there are really two pertinent questions. One question is: do animals have consciousness? But another is: do animals have mental states, whether or not conscious? And it may be that the distribution of consciousness in the animal kingdom is to some extent limited, while the distribution of mentality is more widespread. So, while it may be the case on HOTT that insects lack consciousness, it remains open that insects may nonetheless have *minds*.

HOTT thus strikes what we regard as the perfect balance: while not ruling out that all nonhuman animals have consciousness, the theory is consistent with many creatures having it. So, while the theory plainly does not face the sufficiency horn of Birch’s dilemma of demandingness, it does not face the necessity horn either. And the theory explains not just the distribution of consciousness, but also the distribution of *types of* consciousness, better than rival theories.

**Acknowledgements and Funding Information**:

We thank Richard Brown, Alex Kiefer, Rocco Gennaro, Joseph Gottlieb, Claudia Passos, David Pereplyotchik, Adriana Renero, David Rosenthal, Dan Shargel, and the audience at the 2023 Science of Consciousness Conference in Taormina, Sicily for their helpful discussions of or comments on this material. This work was supported by a Lycoming College Professional Development Grant.

**Conflict of Interest Statement**

The authors have no conflicts of interest to declare.

**References**

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

Bayne, T., Frohlich, J., Cusack, R., Moser, J, Naci, L. (2023). Consciousness in the cradle: on the

emergence of infant experience. *Trends in Cognitive Sciences* S1364-6613(23), 00214-0.

Barron, A. B. and Klein, C. (2016). What insects can tell us about the origins of consciousness.

*Proceedings of the National Academic of Science* 113, 4900-4908.

Berger, J. & Brown, R. (2021). Conceptualizing consciousness. *Philosophical Psychology* 34(5), 637-659.

Berger, J. & Mylopoulos, M. (2021). Default hypotheses in the study of perception: A reply to

Phillips. *Journal of Consciousness Studies*28(3-4), 206-219.

Beran, M. J. (2019). Animal metacognition: A decade of progress, problems, and the development of

new prospects. *Animal Behavior and Cognition* 6(4), 223-229.

Birch J. (2022) The search for invertebrate consciousness. *Noûs* 56(1), 133-153.

Birch, J., Ginsburg, S., & Jablonka, E. (2020). Unlimited associative learning and the origins of

consciousness: a primer and some predictions. *Biology & Philosophy* 35, 56.

Block, N. (2007). Consciousness, accessibility and the mesh between psychology and

neuroscience. *Behavioral and Brain Sciences*30(5), 481-548.

Brown, R. (2012). The myth of phenomenological overflow. *Consciousness and Cognition* 21(2), 599-

604.

Brown, R. (2015). The HOROR theory of consciousness. *Philosophical Studies* 172(7), 1783-1794.

Brown, R., LeDoux, J. & Rosenthal, D. (2021). The extra ingredient. *Biology & Philosophy* 36(2), 1-4.

Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends*

*in Cognitive Sciences* 12(5), 187-192.

Carruthers, P. (2000). *Phenomenal Consciousness: A Naturalistic Theory*. Cambridge: Cambridge

University Press.

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

Oxford University Press.

Carruthers, P. & Gennaro, R. (2020). Higher-order theories of consciousness. *The Stanford*

*Encyclopedia of Philosophy*(Fall), E. N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2020/entries/consciousness-higher/>.

Clayton, N., Dickinson, A. (1998). Episodic-like memory during cache recovery by scrub jays. *Nature*

395, 272–274.

Cowey A, & Stoerig P. (1995). Blindsight in monkeys. *Nature* 373(6511), 247-249.

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

York: Viking Press.

DeGrazia, D. (2019). Animal self-awareness: Types, distribution, and ethical significance. In *The*

*Routledge Handbook of Animal Ethics*. New York: Routledge.

Descartes, R. (1988). *Descartes: Selected Philosophical Writings*. J. Cottingham, R. Stoothoff,

and D. Murdoch (trans.). Cambridge, UK: Cambridge University Press.

Drayson, Z. (2012) The uses and abuses of the personal/subpersonal distinction. *Philosophical*

*Perspectives* 26(1), 1–18.

Dretske, F. (1995). *Naturalizing the Mind*. Cambridge, MA: MIT Press.

Gennaro, R. J. (2012). *The Consciousness Paradox: Consciousness, Concepts, and Higher-Order Thoughts*.

Cambridge, MA: MIT Press.

Gennaro, R. J. (2018). Animal consciousness. In *Encyclopedia of Animal Cognition and Behavior*.

New York: Springer, pp. 1-14.

Halina, M., Harrison, D., & Klein, C. (2022). Evolutionary transition markers and the origins of

consciousness. *Journal of Consciousness Studies* 29(3-4), 62-77.

Hare, B., Call, J., Agnetta, B., & Tomasello, M. (2000). Chimpanzees know what conspecifics do and

do not see. *Animal Behaviour* 59(4), 771–785.

Harman, G. (1987), (Non-solipsistic) conceptual role semantics. In *New Directions in Semantics*,

E. Lepore (ed.), London: Academic Press, pp. 55-81.

Jamieson, D. & Bekoff, M., 1992. Carruthers on non-conscious experience. *Analysis*, *52*, 23-28.

Koch, C., Massimini, M., Boly, M., & Tononi, G. 2016. Neural correlates of consciousness: progress

and problems. *Nature Reviews Neuroscience,* *17,*307-321.

Kohda, M., Bshary, R., Kubo, N., Awata, S., Sowersby, W., Kawasaka, K., Kobayashi, T., Sogawa, S.

(2023). Cleaner fish recognize self in a mirror via self-face recognition like humans. *Proceedings of the National Academic of Science* 120(7), e2208420120.

Krachun, C., Carpenter, M., Call, J., & Tomasello, M. (2009). A competitive nonverbal false belief

task for children and apes. *Developmental Science*, 12(4), 521–535.

Lau, H. (2022). *In Consciousness We Trust: The Cognitive Neuroscience of Subjective Experience.* Oxford:

Oxford University Press.

Lau, H., & Rosenthal, D. M. (2011). Empirical support for higher-order theories of conscious

awareness. *Trends in Cognitive Sciences* 15(8), 365–373.

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

*Brains*. New York: Viking Press.

LeDoux, J., Birch, J,. Andrews, K., Clayton, N. S., Daw, N.D., Frith, C., Lau, H., Peters, M. A.

K., Schneider, S., Seth, A., Suddendorf, T., Vandekerckhove, M. M. P. (2023). Consciousness beyond the human case. *Current Biology* 33(16), R832-R840

Lycan, W. G. (1996). *Consciousness and Experience*. Cambridge, MA: MIT Press.

Margolis, E. & Laurence, S. (2015). *The Conceptual Mind: New Directions in the Study of Concepts*,

Cambridge, MA: MIT Press.

McDowell, J. 1994. *Mind and world*. Harvard University Press.

Michel, M. & Morales, J. (2019). Minority reports: Consciousness and the prefrontal cortex. *Mind &*

*Language* 35(4), 493-513.

Merker, B. H. 2007. Consciousness without a cerebral cortex: a challenge for neuroscience and

medicine. *Behavioral and Brain Sciences* *30*, 63-81.

Nagel, T. (1974). What is it like to be a bat? *Philosophical Review* 83(4), 435-450.

Neander, K. (2006). Naturalistic theories of reference. In *The* *Blackwell Guide to the Philosophy of*

*Language*, M. Devitt & R. Hanley (eds.). Malden, MA: Basil Blackwell, pp. 374–391.

Newen, A. & Bartels, A. (2007). Animal minds and the possession of concepts. *Philosophical*

*Psychology* 20(3), 283-308.

Passos-Ferreira, C. (2023). Are infants conscious? *Philosophical Perspectives*, 1-22.

Phillips, I. 2021. Scepticism about unconscious perception is the default hypothesis. *Journal of*

*Consciousness Studies, 28*(3-4), 186-205.

Plotnik, J. M., Waal, F., & Reiss, D. (2006). Self-Recognition in an Asian Elephant. *Proceedings of the*

*National Academy of Sciences* 103, 17053–17057.

Quine, W. V. (1960). *Word and Object*. Cambridge, MA: MIT Press.

Rosenthal, D. M. (1991). The independence of consciousness and sensory quality. *Philosophical Issues*

1, 15–36.

Rosenthal, D. M. (2005). *Consciousness and Mind*. Oxford: Clarendon Press.

Rosenthal, D. M. (2008). Consciousness and its function. *Neuropsychologia* 46(3), 829-40.

Sauret, W. & Lycan, W. G. (2014). Attention and internal monitoring: A Farewell to HOP. *Analysis*

74 (3), 363-370.

Seth, A. K. and Bayne, T. (2022). Theories of consciousness. *Nature Reviews Neuroscience* 23(7), 439-

452.

Shepherd, J. & Mylopoulos, M. (2021). Unconscious perception and central coordinating agency.

*Philosophical Studies,* 178(12), 3869-3893.

Tononi, G. (2012). *Phi: A Voyage from the Brain to the Soul*. New York: Pantheon.

Tononi, G., & Koch, C. (2015). Consciousness: here, there and everywhere? *Philosophical Transactions*

*of the Royal Society B* 370: 20140167.

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

University Press.

Whiten, A. (2013). Humans are not alone in computing how others see the world. Animal Behaviour

86(2)*,* 213–221.

Wild, M. (2021). Conceptual thought without language? The case from animal cognition. In *Concepts*

*in Thought, Action, and Emotion: New Essays*, C. Demmerling & D. Schröder (eds.). New York:

Routledge, pp. 99–116

Yoshida M & Isa T. (2015). Signal detection analysis of blindsight in monkeys. *Scientific Reports* 5,

10755.

Young, B., Keller, A. & Rosenthal, D. M. (2014). Quality-space theory in olfaction. *Frontiers in*

*Psychology* 5, 1-15.

1. Henceforth we typically drop the ‘nonhuman’ qualifier, and refer to such creatures as ‘animals’. [↑](#footnote-ref-1)
2. On various uses of ‘consciousness’ and related expressions, see, e.g., Berger & Brown, 2021: Section 2. Going forward, uses of such expressions refer to phenomenal consciousness, unless otherwise specified. [↑](#footnote-ref-2)
3. For an overview of such theories, see, e.g., Seth & Bayne, 2022. [↑](#footnote-ref-3)
4. Although the objection is often framed in terms of human infants too, we focus here on the case of nonhuman animals, though what we say here arguably extend to infants as well. [↑](#footnote-ref-4)
5. For a recent defense of HOTT, including replies to many objections, see, e.g., Berger & Brown, 2021. [↑](#footnote-ref-5)
6. We acknowledge that some of the evidence of nonconscious perception has recently been challenged (see, e.g., Phillips, 2021). Some urge that what might seem to be evidence for subliminal perception is instead evidence for either for weakly conscious states that are simply unreported due to stringent criteria for such responses or for nonmental states that are incapable of driving genuine action. But skepticism about nonconscious mentality is questionable for several reasons that we cannot explore here (but see, e.g., Berger & Mylopoulos, 2021). [↑](#footnote-ref-6)
7. Lycan has since given up HOP theory; see, e.g., Sauret and Lycan, 2014. [↑](#footnote-ref-7)
8. Some theories, such as IIT, may also have organism-neutral motivations. But the conditions for consciousness on such theories are thereby typically so undemanding that they face not the necessity, but the sufficiency, horn of the dilemma. On typical interpretations of IIT, for example, the theory is highly undemanding insofar as it attributes consciousness not only to all nonhuman animals, but to anything that exhibits a sufficiently high degree of informational integration, including simple nonliving systems such as logic gates (e.g., Tononi & Koch, 2015). So, while HOTT may not have an advantage over such theories in this connection, it does have an advantage insofar as it does not face the sufficiency horn, as we argue shortly. [↑](#footnote-ref-8)
9. We thank Richard Brown for raising this objection to us. [↑](#footnote-ref-9)
10. Our account nonetheless differs from Gennaro’s in certain ways. Perhaps most saliently, Gennaro defends both a version of concept nativism, on which some concepts are innate, and a form of conceptualism, on which experiential states such as perceptual states are, like ordinary thoughts, conceptual states too. But we remain neutral here regarding both nativism and conceptualism. Notice, however, that HOTT does not presuppose either. Though, on HOTT, HOTs are conceptual as well as necessary and sufficient for consciousness, HOTs are themselves also typically theorized to be seldom conscious (see, e.g., Rosenthal 2005, p. 9). Rather, they are the states in virtue of which perceptual or other states are conscious. HOTT is thus compatible with perceptual states’ being nonconceptual in some way (for a version of nonconceptualism about perception, see, e.g., Dretske, 1995). It is thus reasonable to think that all concepts might be acquired via a learning process involving perception that is both nonconceptual and nonconscious (for a similar type of account, see, e.g., Rosenthal 2005, pp. 203ff). [↑](#footnote-ref-10)
11. To be clear, we are not assuming an inferentialist or conceptual-role account of the nature of conceptual content on which concepts’ contents are individuated by those concepts’ roles in inference—that is, in terms of their inferential connections to other concepts (e.g., Harman, 1987). Our account of ease of concept acquisition is consistent with virtually any theory of the metaphysics of content, such as varieties of teleosemantics on which a concept’s content depends only on its standing in the appropriate evolutionarily developed tracking relation to what it represents (e.g., Neander, 2006). It may be that concepts are variably difficult to acquire insofar as they require possessing different ranges of concepts, though possessing those different ranges of concepts is what puts a concept into the relevant tracking relation, thereby determining that concept’s content. [↑](#footnote-ref-11)
12. One might think that there is good evidence that nonhumans can be and often are aware of cognitive states—namely, in cases of animal metacognition (see, e.g., Beran, 2019; we thank Robert van Gulick for this objection). But there are several things to say. First, we are not committed to the view that no nonhuman animals possess conscious thoughts. Indeed, we believe HOTT would be even more welcomed if it attributed types of consciousness more widely. Secondly, however, it is not obvious that metacognition involves the deployment of concepts of other mental states. It could be, for example, that the metacognitive assessment of a state as being accurate or inaccurate operates nonconceptually or subpersonally (e.g., Lau, 2022). [↑](#footnote-ref-12)
13. Of course, such a three-fold distinction requires an account of the nature of personal-level mental states independent of consciousness. But many theories within the metaphysics of mind, such as varieties of identity theory or functionalism, are compatible with the individuation of mental states in terms of their neural bases or causal roles, which can occur outside of consciousness. [↑](#footnote-ref-13)