

Theories of Consciousness and the Evolutionary Origins of Consciousness

Morgan Clouser
Lycoming College

There is nothing seemingly more interesting to us as humans than our own experiences. We think, desire, feel, and understand, and there is something that it is like to experience these mental states. There is something that it is like to smell flowers and there is something that it is like to hear music. This is consciousness. A mental state is conscious when there is something that it is like to be in it (Nagel, 1974). But what is consciousness in scientific terms? And when did species develop consciousness in evolutionary history? These issues have long been debated within many disciplines including philosophy, archaeology, biology, psychology, and more.

In this essay, I explore an explanation of what consciousness is and what creatures may have it. This will be a theoretical framework for an investigation into archaeological records to find evidence of when consciousness arose in evolutionary history. The archaeological findings may upend contemporary assumptions about when consciousness arose because previous research on the topic has been scarce. I take an interdisciplinary cognitive science approach to this question, drawing on insights from philosophy, cognitive archaeology, and linguistics to form a better understanding of consciousness and its history.

I first clarify in Section I what the issue is in the philosophical debates about consciousness and discuss how cognitive archaeologists understand these issues. Then, in Section II, I explain why I choose a theory of consciousness to explore its implications, the higher-order thought (HOT) theory (e.g., Rosenthal 2005; Weisberg 2011). Section III describes the basics of HOT theory and

addresses a few challenges that HOT theory faces in regard to finding evidence for consciousness in the material record. In Section IV, I describe the perspective of Joseph LeDoux (2019), a HOT theorist, regarding when consciousness arose and raise some challenges to his arguments. Regarding the archaeological issues, I then create in Section V a model for what we are to expect in the archaeological record as evidence of HOTs. Finally, I compare the expectations of this theoretical model to already understood archaeological lines of evidence to render a full analysis of the expectations of HOT theory regarding the evolution of consciousness. These findings may thus guide cognitive archaeology and other studies of the origins of consciousness in the future.

I. Philosophical Preliminaries

As noted, many following Nagel (1974) describe states of consciousness as those for which there is “something that it is like” to have them or that have a feeling or qualitative aspect. But “consciousness” and related expressions are often used in many ways. To clarify my discussion, then, it is important to make a distinction between different kinds of consciousness.

Philosophers working on consciousness often distinguish at least three different uses of “conscious” or types of consciousness (Rosenthal, 1993). To begin with, we often describe creatures as conscious, which we may describe this way:

Creature consciousness: a creature C is creature conscious just in case C is in a state of being awake and responsive to stimuli, as opposed to being asleep and unresponsive to stimuli.

Creature consciousness is a general property of creatures by virtue of whether they may be awake and receptive to their environments or not. It seems obvious that many nonhuman animals that evolved millions of years ago were capable of being awake or asleep. Sleep cycles arose very early in evolutionary history with simple-celled organisms before brains for metabolic reasons, at least 1 billion years ago. Researchers have shown that even hydra, very small aquatic organisms without brains, enter sleep cycles (Greenwood, 2021). But this kind of consciousness is not often thought to raise deep philosophical mysteries and so is not the focus of this investigation.

We do, however, also often describe creatures as being conscious of things, which we may describe this way:

Transitive consciousness: a creature C is transitively conscious of X just in case C is in a mental state M that makes C responsive in some way to X.

This kind of consciousness refers to relations between creatures and the mental states they have about other things. For example, we might say that a dog is conscious of the ball in front of it because it can see it. But like creature consciousness, this type of consciousness is not often thought to pose serious philosophical questions and thus is also not the focus here.

Rather, our focus concerns the use of “consciousness” that applies to mental states themselves:

Phenomenal consciousness: a mental state M is phenomenally conscious just in case there is something that there is “like” to be in M, or there is a quality to being in M.

This type of consciousness is a mental state where there is a certain quality of being in it. Phenomenal consciousness is often thought to raise many questions and seems to be a mystery. To determine whether one is phenomenally conscious or not is a difficult task. Indeed, some think it is even possible to imagine people just like us physically and functionally, but that lack phenomenal consciousness—that is, that we can imagine so-called philosophical zombies (e.g., Chalmers 1996). And if that is the case, it might seem that there is no hope to explain consciousness in physical or other scientific terms. The philosopher David Chalmers (1995; 1996) thus calls the mystery of (phenomenal) consciousness “the hard problem” because it seems impossible to understand what an answer to the question would even entail. Explaining what consciousness entails can be contrasted with what Chalmers calls “easy problems,” such as those that involve understanding the physical processes of the brain that can be directly observed. Consciousness does not seem as though it is something that can be directly observed, which explains why it has been so heavily debated.

Although seemingly daunting, I briefly argue here that there is no hard problem to consciousness. I assume a physicalist approach, which holds that consciousness is not separate from our extended bodies and is a physical part of us or our physical functioning. This is because recently many theorists have developed different theories of consciousness that attempt to explain what consciousness could be.

Such theories often begin with the observation that not only is there evidence of phenomenally conscious mental states such as conscious perception, but there is also evidence of unconscious perception as well. Although many thinkers, at least going back to Descartes (1641), have assumed that all mentality and mental states are conscious, a rich literature has developed studying unconscious mentality, through investigation of dreaming, anesthesia, implicit bias, blindsight, and more (Baars, 2005). Unconscious perception, for example, has been demonstrated through experiments such as those involving visual masking and blindsight. Blindsight is a phenomenon that can occur in certain patients that may have had damage to their visual cortex in the brain, losing their conscious perception (Kletenik et al., 2021). If in the case that a researcher asks a patient, for example, to point to a visual stimulus, the patient will say that they cannot because they cannot see it. However, if the researcher asks them to try, more than 80% of the time the patient will accurately point to the visual stimulus. A natural explanation of this kind of finding is that people see, but do not consciously see, the visual stimuli. This is just one example of how unconscious perception can occur.

Theories of consciousness attempt to explain the difference between conscious and unconscious perception. For example, neural theories of consciousness explain the difference by positing the relevant neural correlates of consciousness (Metzinger, 2000), while higher-order theories claim that a specific type of higher-order awareness is needed to be in a conscious state (Rosenthal, 2005). But there are many such theories (for review, see, e.g., Seth & Bayne 2022).

What theory one endorses will naturally cause very different predictions about not only which creatures have consciousness, but also for when consciousness might have evolved. And while there has been much debate in cognitive science over what kinds of

creatures have consciousness (e.g., Rosenthal, 2008; LeDoux, 2019; Godfrey-Smith, 2020; Birch, 2022), there has been comparatively little debate on when consciousness arose in the evolutionary record.

There is a reason to ask when consciousness emerged in history. Since much mentality is shown to occur unconsciously, then there is a question to ask why any mentality is conscious. Plainly, humans have phenomenal consciousness. But what could explain the emergence of conscious mentality? If it is thought that single-celled organisms might have consciousness, then it must have arisen very early on in evolutionary history. If it is thought that only humans enjoy consciousness, then it must have developed incredibly late in history. LeDoux (2019) has previously claimed that only humans have phenomenal consciousness. Others, such as Godfrey-Smith (2020), claim that more creatures such as invertebrates are phenomenally conscious, thus suggesting that phenomenal consciousness might have evolved around the same time as the early arrival of creature consciousness.

In archaeology, mental states are often viewed in a vastly different way than in philosophy. Archaeologists do not distinguish between mentality and consciousness as philosophers do. Questions about mentality differ from questions about consciousness because arguably consciousness is a type of mentality. The main goal of the branch of archaeology called cognitive archaeology is to study the minds of humanity in the past and how mentality affects behavior. The focus of this area of study is to discern how the mind works which affects decisions made in the past to explain behavior. Within cognitive archaeology, there has been a large debate between two views of the mind (Abramiuk, 2012). The view of archaeologists like Binford (1973) regard the mind as rational and universal, meaning the mind functions by making decisions most beneficial to itself and that this occurs across all of humanity at the same time. Decisions are made based on the best cost-benefit ratio. In this case, cognitive abilities are universal across all humanity because of neurophysiological structures in the human brain. On this view, mentality and cognitive abilities would arise with relevant brain structures and can be understood through observations.

On the other hand, archaeologists like Bordes (1968) think that the mind is empirical and relative, meaning the mind reacts individually and varies across societies and cultures. Behavior and

cognitive abilities are a result of individuals' surroundings from social and cultural ideals. All humans have their own beliefs that stem from their own cultures. Minds differ among the beliefs they hold, which are created by social interactions. On this view, the question of when mentality arose is not due to when physical brain structures arose, but when certain types of beliefs arose that differ from others. Since beliefs are a certain kind of mental state and consciousness is a certain kind of mental state, humanity may differ in the kinds of mental states one is having. Across all humanity, everyone may not have been simultaneously conscious.

The large difference between the archaeological schools of thought in understanding how the mind works results in different outcomes when using a theory of mind to study the past. But since cognitive archaeologists typically do not distinguish conscious and unconscious mentality, the question of when consciousness specifically arose, as opposed to mere mentality, has not been clearly investigated. In any case, regardless of one's theoretical background, it is not obvious that consciousness is something that everyone automatically has. Archaeology and philosophy must come together to create a better understanding of the mind and its evolution. In the next section, I discuss why we must use a philosophical theory of consciousness to help explore the minds of the past.

II. Approaches to Studying the Evolutionary Origins of Consciousness

In this section, I will draw upon Birch's (2022) distinctions between types of approaches to what creatures may have consciousness. I use these distinctions to decide what approach I will use in order to explore my question of when consciousness might have arisen in history.

Birch introduced three different types of methodologies or approaches to settle issues about what kinds of creatures have consciousness. These approaches include what he labels as theory-neutral, theory-heavy, and theory-light. These different kinds of approaches can be used to determine how much theory, if any, should be used when attempting to settle the issue.

First, Birch introduces what he calls the theory-heavy approach. This approach commits to using one developed theory of human consciousness and then applying that same theory to animals

to see if they also meet the requirements of consciousness or not. For example, we can commit to a theory such as the global-workspace theory of consciousness in humans, which holds that mental states become conscious when representations become widely available in what is called the global workspace realized by the prefrontal cortex of the brain (Baars, 2005). Using this theory, animals could be in conscious states if they share the same neural features of the global workspace that humans have.

In order to use this approach to answer when consciousness arose in history, one can choose one theory of consciousness in humans and then look for evidence of consciousness as understood by that theory in the archaeological record. For example, if one applies a global workspace theory of consciousness, they would look for brain functions in the archaeological record that would be sufficient for an integrated global workspace, which is required for consciousness under this theory (Baars, 2005). This can be viewed by looking at the dynamic functioning of the integrated parts of the brain.

The second approach, the theory-neutral approach, considers theories of consciousness to be too speculative to simply choose one when predicting what creatures have consciousness. To avoid choosing a specific theory, theory-neutral approaches must look at many different types of research to understand the minimum requirements for consciousness that are prevalent in each theory. To use this theory in this research, one would need to take a very generalized idea of what consciousness is without any theoretical assumptions and then look for evidence of that, which would be difficult.

Lastly, Birch's theory-light approach falls in the middle of the previous two. The theory-light approach does not fully commit or deny theory altogether. Instead, it uses a broader hypothesis that is compatible with a large range of theories of consciousness. For his work, Birch uses what he calls the facilitation hypothesis as his broader hypothesis that is compatible with many theories. The facilitation hypothesis states that "phenomenally conscious perception of a stimulus facilitates, relative to unconscious perception, a cluster of cognitive abilities in relation to that stimulus" (Birch, 2022, p. 133). This hypothesis does not commit to any one theory of human consciousness, but Birch claims that most modern

theories of human consciousness are compatible with it, including global-workspace theory, higher-order theories, integrated-information theory, and more. To use this approach, one would have to agree that the facilitation hypothesis is compatible with all current theories of human consciousness.

However, each of these approaches have their own pitfalls. Birch himself claims that it is impossible to avoid theory completely as the theory-neutral approach attempts to do. To meet any minimum requirements for consciousness, there must be some theory involved to understand behaviors and how these may be relevant to consciousness, as Birch believes they are. Even if one does not fully commit to a theory of consciousness, there is still some kind of theoretical assumption being made about consciousness, like the facilitation hypothesis. Some kind of theoretical assumption is made when thinking about consciousness.

The theory-light approach is also problematic. As Birch shows, the facilitation hypothesis only supports theories of consciousness that deem that consciousness has a function. Functions of consciousness are those that facilitate certain cognitive abilities or help a creature to survive. This leaves out theories that do not think that consciousness necessarily has any function, such as David Rosenthal's (2008) promising version of HOT theory. On this view, much, if not all, complex behavior could occur without a creature's being in conscious states. A behavior, such as grabbing an umbrella because it is raining outside, would thus not be evidence that a person has conscious thoughts or desires. It may be that such behavior is caused by the unconscious first-order thought that it is raining, even if the individual is unaware that they think it is raining outside.

This leaves the theory-heavy approach. Birch's own criticisms of this approach are that applying too strong of a sufficient condition for consciousness in humans may not be reasonable when applying the theory to animal consciousness (Halina, et. al., 2022). For example, even if global-workspace theory were the correct theory of human consciousness, it is not obvious that it must be the correct theory of nonhuman consciousness. Birch thus ultimately believes that animals may have the same type of consciousness that humans have, and a theory must cover all creatures, not just humans.

However, the theory-heavy approach seems to be the only possible way to look for evidence of consciousness in the evolutionary record. There needs to be something to look for when searching in history for consciousness. The theory-neutral approach is nearly impossible to achieve, and the theory-light approach leaves out promising theories. I thus move forward with a theory-heavy approach. The question remains, however: which theory ought we to adopt to explore its implications for the evolutionary origins of consciousness? In the next section, I argue that we can and should pick a standard version of the HOT theory.

III. The Basics of HOT Theory

In his work, Birch does not seem to think that higher-order thought (HOT) theory is a particularly promising account of consciousness, so he does not discuss it in detail. However, higher-order theories of consciousness have become an increasingly popular type of theory in understanding what it means to be in a conscious psychological state. Within these theories, to be conscious is to have what is called a higher-order (HO) state about one's psychological state.

The main motivation for such HO theories is what Rosenthal (1997) has called the Transitivity Principle:

Transitivity Principle (TP): A conscious state M is a state whose subject is, in some way, aware of being in M.

Psychological states, like being hungry or feeling a certain emotion, are first-order (FO) states. These states are FO insofar as they are about the world, are not mental states, and are not conscious in themselves. One does not have to be aware that they are in these states, because the TP shows the reason for these psychological states' occurring unconsciously. If one is in a state, but in no way aware that one is in it, then that state is not conscious. But this is logically equivalent to claiming that a mental state is conscious only if one is somehow aware of being in it—that is, if one has a HO state, which is HO insofar as it is directed at or about another mental state. HO theories attempt to explain this commonsense claim but disagree about the mechanism that enables us to be aware of our FO mental states.

According to a common type of HO theory, higher-order thoughts (HOTs), which are an ordinary type of thought, make one aware of their current states:

HOT theory: A mental state M is conscious just in case one is aware of being in M via a suitable HOT.

In such a view, for example, one can be hungry, but when they become aware of their state of hunger via a suitable HOT, that is when the state of hunger becomes conscious. Roughly, to be in a conscious state of hunger, one must think in a suitable way “I am hungry.” In most versions of HOT theory, such a HOT is distinct from the FO psychological state that it is aware of.

When searching for evidence of consciousness, HOT theory is particularly compelling. HO theories are the only type of theories of consciousness that try to explain the Transitivity Principle. But there are many other reasons to be drawn to HOT theory. For example, one’s experiences can differ when learning new things from perception, called perceptual learning. Learning new things, like new words for certain experiences, can impact how one perceives the world (Rosenthal, 2008). An example of perceptual learning can be seen in wine tasting. If one is inexperienced in tasting wines, they might not be able to tell the difference between cheap and expensive wines and would not be able to pick out specific flavor profiles. The more experienced one is in wine tasting, the more they can taste the differences in flavors and cannot enjoy the taste of the cheaper wines like they once did. Learning specific words for noticing small details in wine flavors permits one to make distinctions that they might not have been able to make previously because they did not know the words.

This phenomenon occurs because one is learning whole new concepts for tastes. One’s experience is categorically different when they use different words for tastes. The best explanation for this is HOTs. To have an experience is to be aware of what one is perceiving. Experiences can differ when learning new things from perception. Having an acquired taste for wine is having a learned taste, which differs by having HOTs about such a thing. Other theories of consciousness, like global-workspace theories, are unable to describe this phenomenon of perceptual learning.

Additionally, there is empirical evidence that supports HOT theories (Lau and Rosenthal, 2011). Using visual masking with matched task performance, Lau and Rosenthal (2011) were able to determine that conscious brain activity is associated with the dorsolateral prefrontal cortex (dlPFC). This experiment with matched task performance eliminated the global workspace theory which claims that brain activity in the dlPFC is associated with conscious awareness and task performance. Task performance was matched, but visual consciousness still shows activity in the dlPFC, which supports higher-order theories.

The best evidence of consciousness under HOT theory would be a verbal report that clearly indicates that one is in particular mental states, thereby expressing their HO awareness of those states. One could not report being in a state that is unconscious. Most theories regard verbal report as the gold standard evidence for consciousness. It is easy to know there was consciousness when there is a report of somebody being in that mental state. This awareness could be indicated in speech or in concise writing. For example, if someone wrote, “I want food,” then this would be evidence that they are aware of their desire for food and are conscious. HOT theory has a natural explanation of why verbal report is the gold standard. Such a report is, according to the theory, also evidence that one has the relevant sort of higher-order mental state—that they are aware of their desires (Rosenthal, 2008).

However, a puzzle for HOT theory in particular and consciousness studies more generally is this: if verbal report is the best or only evidence that there is consciousness, then we cannot know if nonhuman animals or even ancient humans had consciousness because we cannot obtain any evidence of them being able to report their mental states. LeDoux (2019), a HOT theorist, thus claims that since animals and past humans cannot verbally report, there is no other way that we can search for consciousness within them. Thus, he concludes that since only *homo sapiens* have the capability for language, consciousness evolved with them.

Since we cannot directly observe somebody’s mentality, this becomes an even greater challenge for archaeology. Archaeology studies the past through material remains. If there are no material correlates to study for evidence of consciousness, then it would be

impossible to move forward. We must thus find other indirect lines of evidence for HOTs.

According to HOT theory, it is unclear what evidence of consciousness could be other than verbal report. Different theories show that consciousness may have a function that could be used as evidence for it instead and would not require verbal report. For example, the global-workspace theory holds that one is conscious if information in one area of the brain is sufficiently available to other areas of the brain. This kind of information integration would be caused by ideas of the prefrontal cortex lobe of the brain. Evidence of consciousness could thus be evidence of complex planning behavior that would require states in the global workspace. According to Pain (2022), stone tool creation takes complex intentional action, which is from information processing in the mind. Under global-workspace theory, consciousness was necessary for complex planning behaviors like stone tool creation, so material remains of stone tool creation would show that there was consciousness. This period of consciousness development would be even later than what other HOT theorists such as LeDoux posit, but at least this could be other evidence sufficient for consciousness that is not solely verbal report.

However, again, in a standard version of HOT theory, it is doubtful that consciousness holds any utility or function (Rosenthal, 2008). Any executive functioning that mental states could do for us need not occur with accompanying HO states. Interactions between FO states can occur, like in beliefs and desires, without HO states interfering or even occurring at all. If consciousness does not hold much utility and does not add function to human behavior, it is hard to find any other evidence for it. What may seem like evidence of consciousness would look no different than evidence of any other mental activity occurring without consciousness. If we found evidence of complex planning behavior, like stone tool production mentioned above, this behavior could be the result of unconscious FO mental states alone.

Since Rosenthal (2008) claims that consciousness has no executive function or benefit to thinking and reasoning, there still is a question of why consciousness arose at all if there is no function associated with it. Pain (2022) rejects the idea that producing syntax, or language production, was a result of a random genetic mutation.

If it were, these capacities would be restricted to only humans somewhere around 100,000 years ago. Pain's claim on why humans have reached a sophisticated level of language production was the result of syntactical capacities increasing due to the effects of tool production and language production that developed in the evolution into *homo sapiens*. Having syntactic abilities is not sufficient for the ability for language production because syntactic abilities can occur in any species that has intentional action, so rather they are necessary for language production.

Likewise, Rosenthal hypothesizes that consciousness was an evolutionary spandrel that occurred with the beginnings of language. Producing language uses many abilities in cognitive systems, like complex planning, intentional action, and memory. These individually play a role in producing language but are not sufficient for it, explaining why animals and other creatures may have these abilities but do not have language. As language use increased, people were noticing that they were talking. The sum of humans' cognitive abilities allows for a sophisticated inference from their own behaviors to their internal states that cause behavior. It is likely that humans then developed the capacity to have concepts of their internal states to form HOTs about the causes of their speech and the behaviors their speech portrayed. Language was necessary for thoughts and desires to become conscious as humans began to observe their own behaviors. One may be conscious of their thoughts and desires by observation or inference of their behaviors. Therefore, consciousness would have arisen even later in evolutionary history than simply the beginnings of *homo sapiens* from 300,000 years ago (Galway-Witham & Stringer, 2018).

To many researchers, the idea that consciousness arose in humans is startling. Many theorists have denied it, such as Tye (2016) and Godfrey-Smith (2020), who claim that crabs and octopi also have it. This disagreement among theorists may partly be because different thinkers endorse different theories and approaches to studying consciousness. In the next section, however, I aim to show why there might not be as much disagreement as there seems.

IV. HOT Theory and the Evolution of Consciousness

One recalls that LeDoux (2019) makes a different prediction about when consciousness arose in history with a theory-heavy HOT

approach. LeDoux claims that, if HOT theory is true and verbal report is the only evidence we have of consciousness, then the only evidence of consciousness that we can have regards humans. He believes that humans had consciousness with the birth of modern *homo sapiens* and they are the only creatures that have it because humans are the only ones with the necessary brain regions for HOTs, likely to be found in the dorsolateral prefrontal cortex (dlPFC). Since humans are the only creatures available to give a verbal report due to language and are the only ones with a developed prefrontal cortex, then humans are the only creatures with HOTs and consciousness. Again, this answer is startling to many theorists.

I argue now that there is at least one way to reconcile what may seem to be this deep debate among consciousness researchers. Rosenthal makes an important distinction between types of consciousness that LeDoux did not consider in his predictions. Verbal report is often considered evidence for what we might call “cognitive consciousness,” but not necessarily for what we might call “qualitative consciousness.” The difference between these types of consciousness involves what kind of FO mental states a creature may have HOTs about. Qualitative psychological states include perceptions and sensations, while cognitive, or non-qualitative, states include thoughts, desires, and intentions.

According to Rosenthal, common sense holds that many nonhuman animals such as dogs have qualitative but not cognitive consciousness. There is reason to think that dogs can feel pain consciously, but there is little reason to think that dogs may have conscious thoughts and desires about the world. Since dogs cannot give a verbal report, there are no experimental findings on dogs having cognitive consciousness. However, there are experimental findings concerning conscious perception in other mammals. For example, it is possible to induce blindsight in monkeys (Yoshida & Isa, 2015), showing that at least some animals can be induced into unconsciously perceiving that which may be ordinarily conscious.

This crucial difference in qualitative and cognitive consciousness is often overlooked by theories of consciousness in humans and other creatures. But given this distinction, it is easy to see why some debates are not really debates at all. HOT theorists can happily grant that many nonhuman animals have qualitative consciousness, while still maintaining that cognitive consciousness

may have evolved rather late in history, and perhaps only in humans. Therefore, I will only focus on cognitive consciousness and the question of when it arose in evolutionary history.

Given that the most promising theory, HOT theory, and current evidence suggest that cognitive consciousness only occurs in humans, it might seem that it arose at the same time that *homo sapiens* evolved. However, I argue that we still must further distinguish between having the capacity for cognitive consciousness and actually having it. Knowing the difference between having a capacity for consciousness and having it will yield very different results about when consciousness arose in history.

LeDoux seems to assume that having the relevant neural structures for HOTs indicates that humans automatically have consciousness. This is a common assumption in cognitive archaeology as well, in that evidence of the relevant neural structures is necessary and that sufficient evidence of certain mental functions occurred (Binford, 1973). But there is no reason to assume that because humans may have always had a capacity for a mental function, such as consciousness, that they actually exercised or developed that function.

This is especially true if HOT theory is correct. After all, HOTs are a type of ordinary thought and just because people can have that type of thought, it does not mean that they necessarily do. For example, all humans have the capacity to think about hard mathematical topics, such as calculus. But just because anyone can have thoughts about calculus does not mean that we do have them or that having such thoughts is necessary to us as humans. So, it is not obvious that having language capacity or the right neurophysiological structures shows that humans had cognitive consciousness.

The latest possible concrete evidence that we could find for such consciousness would come in the form of a verbal report of mentality through language. Today, the Sumerian language of southern Mesopotamia is widely regarded as the first known written language, emerging around 5,000 years ago (Kenanidis, 2013). Extensive research into the language could possibly reveal verbal expressions of consciousness earlier in history. Verbal expressions of consciousness would need to be examples of not simply expressing thoughts or perceptions, but expressing awareness that one thinks or

perceives their own thoughts. An example could be a statement such as “I am hungry.” This would show that one is aware of their own desires and thoughts they have. An expression of simple hunger could look like “Get food now.” There may have been a period before consciousness arose when people simply had FO states of hunger that they could verbally express without being aware that they were hungry and thus were having HOTS about such states.

Although it is not guaranteed that Sumerians were conscious when writing their language around 5,000 years ago, this timeframe still feels incredibly late in history for the arrival of cognitive consciousness. For simplicity, this project will not regard the hunt for the first verbal expressions of consciousness through past literature. To move forward, we must explore whether or not there could be nonverbal expressions of consciousness that occur before written language that would provide evidence of the relevant kind of HO awareness.

V. Exploring the Archaeological Record

If HOT theory is true, what nonverbal, behavioral, and thus material expressions of cognitive consciousness would we expect to see in the archaeological record? These are the implications when people express their HO awareness of their mental lives. I now offer a hypothesis regarding the material correlates of HOTS. If sound, this theoretical model could then be used and applied to already known archaeological cases to determine when consciousness evolved.

In the HOT theory that Rosenthal (2008) develops, consciousness has no utility or function, so there would be no individual material or behavioral correlates to having it. Any immediate action that is driven by mental states that are conscious could arguably be driven by those same states unconsciously. The only evidence we could have for consciousness, then, would be verbal reports in the form of written language.

However, I claim that HOT theory is still compatible with a version of consciousness with some group utility. Consciousness may not have any direct function for the individual, but the possession of it may lead to social consequences that have implications in the material record. The consequences of consciousness may be apparent in groups and societies on a social

level. We can understand this claim through the evaluation of cultural evolution theories of prosocial behaviors (e.g., Norenzayan & Gervais, 2011).

One such prosocial behavior that has a large role in culture is the introduction of religion and religious behaviors (Norenzayan & Gervais, 2011). Religion is a cultural byproduct of promoted prosocial behaviors and facilitates the benefit of others in group settings (Norenzayan et. al., 2016). Religion has four features that have remained stable across all time periods and cultures, and it is commonly believed among cognitive archaeologists that one necessary feature of religion is the individual's ability to perceive other minds and infer the thoughts of others (Norenzayan & Gervais, 2011). This ability is what facilitates the attribution of desires and beliefs to gods because they are humanlike and recognizable. Additionally, religious beliefs are a byproduct of the concern of individuals' reputations socially. An individual within a group becomes aware of their own place and status within their group and the place of others, resulting in religious behaviors that promote the welfare and cooperation of shared ideas in a group to make stronger bonds within it. Religious beliefs have inherent cooperative intentions, facilitating the rise and stability of larger communities (Norenzayan & Gervais, 2011).

This core feature of religion, that an individual is aware of their own thoughts and others' thoughts within a group, is a byproduct of HO awareness. Religion could not be facilitated without this understanding of one another's beliefs and desires—and thus arguably of one's own mental states as well. As such, we can assume that the behavioral and material consequences of the introduction of HO awareness and cognitive consciousness would facilitate religious behaviors that are intended to promote the welfare of a social group. Religion as a consequence of HO awareness would infer that consciousness was necessary for religion to occur, dating back to before religious activity had begun.

In the archaeological record, early human behavior and cognitive abilities are still not well understood. As the topic has become more popular in recent years, more research is being released. We can find accounts of prosocial behaviors that were intended to promote cooperation dating back even before *homo sapiens* to the Neanderthals. We can expect to find accounts of

increased prosocial behaviors regarding contribution to group welfare, before religion, as indicators of the beginnings of consciousness. HO awareness benefited groups by promoting stability and cooperation from shared beliefs. Twomey (2013) claims that prosocial behaviors associated with religion, like future directed cooperation and resolving social dilemmas, can be seen as early as 200,000 years ago. These cognitive abilities can be implied through behaviors related to controlled fire creation (Twomey, 2013). To control fire, not just use fire, humans needed a complex variety of cognitive skills. Controlling fire requires intensive planning, group-level cooperation, intentionality, and social awareness (Twomey, 2013). Although evidence of religious behaviors is currently scarce when studying ancient humans, we can view similar prosocial behaviors in other actions, like controlling fire. These types of actions are sufficient behavioral and material correlates for the evidence of introduced cognitive consciousness within species.

Even with this claim, is there a way to imagine religious and prosocial behaviors without necessarily having HOTs? Some may claim that it is imaginable. We can think of a group of people who are not conscious, but still can have religious thoughts and perform religious behaviors. Why would HOTs be necessary if we could imagine this type of scenario?

But why would there be religious behaviors at all if there was no purpose for them? Religion is a part of culture, which provides important social benefits. These behaviors must have occurred as a response to something, which I claim was the introduction of consciousness to the human species. Without consciousness and HOTs, there would be no purpose to even partake in religious behaviors, as humanity would still only be focused on survival. Consciousness adds a certain quality to life and enhances it in social ways.

VI. Conclusions

We now have a much clearer understanding of the issue at hand concerning when consciousness may have arisen in evolutionary history. We can say with confidence that humans were conscious when they could verbally report their higher-order mental states. This may have occurred late in evolutionary history, around 5,000 years ago. If we had accepted previous predictions like

LeDoux's, consciousness would have arisen with the beginnings of *homo sapiens*, around 300,000 years ago because of neurophysiological changes specific to the species.

In general, there is continually much debate around the origins of consciousness. This paper helps to identify these problems concerning consciousness and further claims that we can view the behavioral implications of consciousness through prosocial behaviors around 200,000 years ago. We have only begun to explore the implications of Higher-Order-Thought theory of consciousness through the archaeological record. This paper will start the long journey of this task at hand.

References

- Baars, B. (2005). "Global Workspace Theory of Consciousness: Toward a Cognitive Neuroscience of Human Experience." *Progress in Brain Research* 150: 45-53.
- Birch, Jonathan (2022). "The Search for Invertebrate Consciousness." *Nous* 56: 133-153.
- Birch, Jonathan (2021). "The Hatching of Consciousness." *Springer* 43: 121.
- Descartes, R. (1641). "Meditations on First Philosophy." In J. Cottingham et al., (Eds.), *The Philosophical Writings of Descartes*, Vol. II. Cambridge: Cambridge University Press.
- Galway-Witham, J. & Stringer, C. (2018). "How did Homo Sapiens Evolve?" *Science* 360: 1296-1298.
- Greenwood, Veronique (2021). "Sleep Evolved Before Brains. Hydras Are Living Proof." *Quantamagazine*.
- Godfrey-Smith, Peter (2020). *Metazoa: Animal Minds and the Birth of Consciousness*. London: William Collins.
- Halina, M., Harrison, D., and Klein, C. (2022). "Evolutionary Transition Markers and the Origins of Consciousness." *Journal of Consciousness Studies* 29, no. 3-4: 62-77.
- Kletenik, I., et al. (2022). "Network Localization of Unconscious Visual Perception in Blindsight." *Annals of Neurology* 91: 217-224.
- Laland K, Seed A (2021). "Understanding Human Cognitive Uniqueness." *Annual Review of Psychology* 4: 689-716.
- LeDoux, Joseph (2019). *The Deep History of Ourselves: The Four-Billion-Year Story of How We Got Conscious Brains*. New York: Viking Press.
- Metzinger, Thomas. (Ed.). (2000). *Neural Correlates of Consciousness: Empirical and Conceptual Questions*. Boston: MIT Press.
- Nagel, Thomas. (1974). "What is it Like to Be a Bat?" *Philosophical Review* 83, no. 4: 435-50.
- Norenzayan, A. and Gervais, W. (2011). *The Cultural Evolution of Religion.* in *Creating Consilience: Integrating Science and the Humanities*. Oxford: Oxford University Press.
- Norenzayan, Ara, et al. (2016). "The Cultural Evolution of Prosocial Religions." *Behavioral and Brain Sciences* 39. Cambridge: Cambridge University Press.

- Pain, Ross (2022). "Stone Tools, Predictive Processing and the Evolution of Language." *Mind & Language* 38, no. 3: 1-21.
- Rosenthal, David (2008). "Consciousness and Its Function." *Neuropsychologia* 46: 829–840.
- Seth, Anil & Bayne, Tim (2022). "Theories of Consciousness." *Nature Reviews Neuroscience* 23: 439-452.
- Twomey, T. (2013). "The Cognitive Implications of Controlled Fire Use by Early Humans." *Cambridge Archaeological Journal* 23 (1), 113-28.
- Tye, Michael (2016). *Tense Bees and Shell-Shocked Crabs: Are Animals Conscious?* New York: Oxford University Press.
- Weisberg, Josh (2011). "Misrepresenting Consciousness." *Philosophical Studies* 154 no. 3: 409-433.
- Yoshida, Masatoshi & Isa, Tadashi (2015). "Signal Detection Analysis of Blindsight in Monkeys." *Scientific Reports* 5.