What Can We Know About Consciousness in Animals?

A Critical Analysis of Claims by Joseph LeDoux

Hilde Andenæs

Thesis

Master of Philosophy

Supervisor: Bjørn Ramberg

Department of Philosophy, Classics, History of Art and Ideas

UNIVERSITY OF OSLO

Spring 2018

Hilde Andenæs

What Can We Know About Consciousness in Animals?

A Critical Analysis of Claims by Joseph LeDoux

CartoonStock.com

©Hilde Andenæs
2018

What Can We Know About Consciousness in Animals?
A Critical Analysis of Claims by Joseph LeDoux
Hilde Andenæs

http//www.duo.uio.no
Print: Allkopi

Abstract

This thesis examines two claims about consciousness in animals by the neuroscientist Joseph LeDoux: 1. Science cannot tell us anything about consciousness in animals, and 2. Animal behavior can be explained in terms of neuroscience. He argues that views that animals are conscious are not based on science. His view stands in contrast to the increasingly accepted view that many animals are conscious. The debate on animal consciousness is complex. Many disciplines are engaged and the debate involves different definitions, terminology, questions and theories. LeDoux raises important points that should be addressed, clarified and responded to.

I discuss challenges with LeDoux's narrow definition of consciousness, which requires explicit knowledge, and argue that his claims about the relation between science and consciousness are inconsistent. If correlates of consciousness can tell us about consciousness in humans then they can also tell us about consciousness in animals. I then examine his second claim. I argue that his neuroscientific theory of consciousness does not warrant exclusion of animal consciousness. His explanation of behavior is not exhaustive. I then identify and argue against two underlying assumptions in his model of the relation between stimuli, consciousness and behavior. I argue that the causal content of consciousness is assessment of behavioral performance in relation to stimuli. Consciousness hence does not directly affect behavior, but indirectly through modification of future behavior. This requires direct access to experience of own behavior, but does not require any kind of explicit knowledge. I discuss reinforcement in light of artificial and natural learning systems, affordances and predictive coding and conclude that consciousness is the discrepancy between expected and actual reinforcement. Consciousness is learning. I propose a new positive marker of consciousness: the ability to modify and learn new behaviors in a changing environment. This generates testable predictions and is a promising direction for the future studies of correlates of consciousness.

Acknowledgements

I started off with the study, and later research, of animal behavior and welfare. As an ethologist I measured behavior. I took for granted that animals have conscious experiences and that behavior can indicate what they feel. When I entered into the philosophy of mind this opened up a world of topics, terminology and challenges related to the phenomenon of consciousness. It was much more complex and far more exciting than I had ever imagined! In this thesis I address what science can tell us about consciousness and I explore the relation between consciousness and behavior. Surprisingly, this relationship turned out to be different than I had anticipated.

I have many animals to thank, both human and non-human:
First, I want to thank my supervisor Bjørn Ramberg for his enthusiastic and positive attitude towards my project. He has contributed with his knowledge and insights to keep my thoughts on track. I would also like to thank all the other teachers I have been fortunate to come across that have shared their thinking, ideas and knowledge.

Thank you to my lovely and awesome family, my husband Frode and our three girls Una, Eva and Susanne, for their support and patience with my obsessive interest and detailed rambling on the topic of consciousness and analysis of behavior.

Finally, I would like to honor all the animals of many different species that I have been privileged to spend time with, both through interaction and observation. I have witnessed the cruelty humans can inflict on animals and the suffering they endure. When animal interests conflict with human interests, many people will justify their actions by referring to the lack of knowledge we have about conscious experience in animals: “We can’t know if they feel anything, so they probably don’t.” I believe this view needs to be properly addressed and challenged.

I have enjoyed close relationships with our many dogs over the years. Especially Fia, with her eagerness for learning and long hikes in the forest, has inspired many thoughts and ideas about how her mind works. I suspect though that our minds and worlds are much further apart than I like to believe. I hope that when I feel she is happy – she really is. I dedicate this thesis to Fia.

Hilde Andenæs
Oslo, 11.06.18

Content

1 Preface

This thesis concerns consciousness in animals. It addresses epistemological and metaphysical issues related to claims against consciousness in animals by the neuroscientist Joseph LeDoux in the book "Anxious. Using the Brain to Understand and Treat Fear and Anxiety" (2015). In November 2017, I attended a multidisciplinary conference on Animal Consciousness hosted by New York University. There were participants from fields such as philosophy, biology, neuroscience and cognitive sciences. Many important topics related to consciousness were discussed and the tendency is towards more species being included in the conscious realm. LeDoux however claims that this is not for scientific reasons. The debates revealed many different definitions, terminology, underlying assumptions and kinds of questions being asked. This makes the debate complex. It is important to get an overview, identify tensions and clarify positions and their relation to science in order to decide how to proceed. I believe a clear response to LeDoux's objections is necessary.

The thesis addresses a broad range of topics related to consciousness and therefore contains ample descriptive material. This material has proved necessary for my understanding and I believe that it may be helpful for the reader to get an overview. The trade-off is that not all topics are discussed exhaustively. Many of them are their own topics of an extensive debate. My main goal has been to get an overview over the debate and to be able to draw some conclusions about correlates of consciousness and how these relate to science. This thesis also challenges some common basic assumptions about the relation between consciousness and behavior and presents an alternative model that suggests a positive marker of consciousness. I hope this thesis can be of interest for people with an interest in animals and consciousness.

The thesis is structured in the following way: The first part contains a general introduction to the phenomenon of consciousness and issues that relate to questions regarding consciousness in animals. Then I present, examine and discuss two statements: 1) Science cannot tell us anything about consciousness in animals, and 2)

Animal behavior can be explained in terms of neuroscience. Each claim is presented separately in their own section followed by a discussion in the following section. Finally I discuss the relation between learning and consciousness, and how a system can evaluate behavioral success, modify and learn new behaviors in a changing environment. I argue that conscious experience can play a role in reinforcement. The ability to modify and learn new behaviors is thus a positive marker of consciousness.

2 Introduction

2.1 The importance of consciousness

Conscious experience is the essence of human life; without consciousness most people do not consider life worth living. We all intuitively know what it is like to experience our own bodily sensations, sensory perceptions, feelings and thoughts, yet we do not know what consciousness is, how it works or if it even has a function. We thus struggle to predict its presence outside ourselves.

Human beliefs about which species belong within the conscious circle have varied over time and within different scientific and religious knowledge- and belief systems. The study of consciousness today has become a multidisciplinary field involving a wide range of subjects such as philosophy, psychology, cognitive science, evolutionary science, neuroscience, physiology, ethology, and computer sciences to mention some. These fields generate many different theories that are based on different assumptions about necessary and sufficient conditions for consciousness to be present (Pereira et al., 2010). The different theories vary in their views on whether or not non-human animals[1] can have conscious experiences. Questions and claims regarding animal consciousness are of great importance and urgency due to the widespread human instrumental use of other organisms.

[1] Animal refers to all organisms that are multicellular eukaryote organisms without cell walls that consume other organic matter. This therefore includes both vertebrates such as mammals, birds, fish, amphibians and reptiles, and non-vertebrates such as insects.

2.2 What is consciousness?

Consciousness is a phenomenon that is hard to pin down with labels in natural human language. It encompasses a myriad of different kinds of experiences, many of which are difficult to verbalize even ineffable. Examples of conscious experiences in humans are: sensory perceptions such as the sound of a violin or the taste of an apple; bodily sensations such as pain, hunger, satisfaction, proprioception, dizziness or the rush of adrenaline in a sudden scary encounter; drives or motivations such as the desire to run; feelings such as happiness, anger, sadness or remorse; and the experience of ones own thoughts in words or other kinds of experiences, such as suddenly understanding or remembering something (Strawson, 1994) or even the feeling that accompanies a statement such as "2+2=5." We all have direct access to these experiences; we know what they are like.

The nature of conscious experience is such that consciousness is difficult to define. Studies and debates are permeated with diverse use of definitions, terms and concepts that are applied in various kinds of questions and explanations. Sometimes different terms will refer to the same thing, yet other times identical terms refer to different things. For example, the term "emotional" can refer to both conscious (Panksepp, 2005) and nonconscious states (Damasio, 2003). Terms that are closely related to consciousness and that are often used interchangeably are: subjective experience, phenomenal experience, awareness, sentience, feelings, emotions, mental states, intentional states and affective states, along with others.

Conscious experience has a subjective viewpoint (Nagel, 1986) that is private and only directly accessible for the subject. Consciousness has an experiential center (Tartaglia, 2016). There may however be organisms that possess non-integrated parts or sensory modalities (Sjølander, 1997; Godfrey-Smith, 2016). A conscious system does not merely detect stimuli; perception has a qualitative feel – a phenomenological aspect, phenomenological property or phenomenal character. As Nagel put it: "..there is something that it is like to *be* that organism – something it is like *for* the organism" (1974, p. 436). In humans, conscious experiences have certain characteristics to the subject: they exist intrinsically, they are composed of phenomenological distinctions, they can be differentiated from each other, they are unified and each experience is definitive in the sense that it includes neither more or less (Tononi & Koch, 2015). Humans have a strong intuition that their conscious experience plays some kind of informing role and in this way plays a causal role for behavior.

Conscious experience can be characterized in terms of levels of knowledge of the experiential center: from anoetic (without knowledge) to noetic (knowledge based) and autonoetic (higher reflective mental) (Vandekerckhove & Panksepp, 2011). There are differing views on at what level the phenomenal aspect enters: rudimentary states at low levels (Vandekerckhove et al., 2014) or at higher levels that involve explicit knowledge (LeDoux, 2015). The transition may be gradual or abrupt. A distinction commonly made is between creature consciousness and mental state consciousness (Manson, 2000). Creature consciousness often refers to an organism being awake, alert and capable of interacting with its environment, as opposed to in deep dreamless sleep, in a coma or anesthetized. Mental state consciousness often refers to a creature having explicit experiences that involve metalevel awareness. The subject is aware of being in the state in the sense that it is aware that the state is occurring to itself and has knowledge of what the state is about (Rosenthal, 1993; Lycan, 2001; LeDoux, 2015). This is also referred to as full-blown consciousness. Some regard intentionality as a mark of the mental (Crane, 1998). The distinction between creature- and mental state consciousness does however not clearly distinguish between a non-phenomenal detection state and a sensory perception, which is an essential and difficult issue.

Conscious is often used in the form "conscious of …" and can thus easily be taken to imply that the subject is conscious *of* something, which can be associated with theories that require higher-order representation (Rosenthal, 2005). Feeling may be considered more direct. "The organism feels pain" may have a different connotation than "the organism is conscious of pain." The latter seems to require that the subject need have explicit knowledge about the state as well as a definition of pain. Another term that captures the phenomenal aspect is sentience: having the capacity to feel (Harnad, 2016). Nonconscious and unconscious are sometimes used interchangeably. A nonconscious process will here refer to a process that is not experienced consciously by the subject. When an organism is unconscious it is in a state where it is not having any conscious experiences.

The phenomenon of consciousness in turn involves other phenomena that have their own conceptual challenges such as cognitive functions like attention, awareness and memory (Lamme, 2003). The study of consciousness is multidisciplinary, and therefore involves participants from different thinking traditions. Philosophers for example, are often concerned with the character and structure of conscious states, and higher order thinking and language. Animal scientists tend to focus on the relation between consciousness and problem-solving. It can be difficult to isolate and compare what exactly consciousness is and does in these different areas. A clear discussion needs to involve clear definitions. Without this it is difficult to disentangle what is factual and what is verbal is these disputes (Chalmers, 2011).

2.3 Questions about consciousness

2.3.1 Proximate and ultimate questions

Consciousness as a phenomenon can be addressed by different questions: *What it is like* to be conscious concerns its phenomenology, what consciousness intrinsically *is* concerns its ontology, how consciousness *works* involves both the proximate explanation with parts and mechanisms and its possible *function* as in how it may be causally related to behavior and ultimately what kind of problem it solves to constitute an adaptive trait in an evolutionary perspective.

The biologist Niko Tinbergen (1963) listed four complementary questions to explain behavior, and these are either proximate or ultimate. These different questions give different kinds of answers. Proximate questions address the proximate mechanisms and processes involved in eliciting the behavior. The proximate question points backwards in time in that it tries to describe all the preceding events that causally contribute to elicit the behavior. This resembles what Dennett (2017) refers to this as a “How come?” question in the sense that it answered by looking at preceding events. A typical proximate question is: How come birds sing in spring? The answer will include all the internal and external factors that causally contribute to the behavior such as temperature and light affecting hormonal levels and the presence of other birds.

Ultimate questions address the evolutionary function of the behavior. This is what Dennett (2017) refers to as the "Why" question. Why does the bird sing in spring? The answer can be to attract mates and to defend its territory, which in turn affects fitness. The answer to this question points forward in time: it says something about what is obtained by the performance of the behavior. This can be viewed as a kind of teleological question, such as Aristotle's fourth question regarding "final cause." This may seem to imply that there is some kind of intelligent design hovering behind. But nature has no plan and no guide; it is a blind and mindless process. Darwin has shown us that there can be design without an Intelligent Designer. Dennett argues that it is legitimate to ask about function (Dennett, 2017). This view may be considered controversial in philosophy, and objected to on metaphysical and ontological grounds. For example, that claiming bacteria do things for a reason is metaphorical, that it can be taken to imply that bacteria have a mind. I will however follow Dennett here. It has great explanatory power in terms of explaining behavior.

2.3.2 The problem of other minds

The starting point for any human observation is the subjective experience of the observer herself. It can be argued that our beliefs about our own experiences are justified by our having the experience, not by a causal link to the experience, nor the mechanisms that form the beliefs (Chalmers, 1996). A mind has direct access to own subjective conscious experience, but only indirect access to the experience of another mind. Questions that ask whether, when or what an organism (or a system) experiences concern "the problem of other-minds." In the debate on animal consciousness it is important to distinguish between metaphysical and epistemological aspects of this cluster of questions. Claims about whether or not we can know *if* other animals are conscious belong to the "Distribution question": Can we know which species are conscious? Claims about *what* other animals can experience belong to the "Phenomenological question": Can we know what those states are like?

Answering "yes" to the distribution question requires criteria for the presence of consciousness. How can we know if a system is conscious? We commonly assume that certain events such as the uttering of words, behavior or neurophysiological activity, correlate with consciousness experience in humans. We use these to investigate consciousness in animals. But we can only deny the human kind of consciousness in animals. It may still be present in forms that are elusive to human sensory and cognitive abilities, or realized by other structures. There are however differing views on whether or not the problem of other species' minds can be solved with a reasonable amount of certainty based on these correlates. On one side there is the view that without direct evidence this is not possible (LeDoux, 2012). On the other side, the view that if consciousness causally affects behavior then it should in principle be possible (Allen & Trestman, 2014). An important question is whether or not the correlates we can observe from other humans differ essentially from the correlates from other species.

The phenomenological question asks whether or not we can have knowledge about the phenomenal content of conscious states in other species. For example, does a certain conscious experience feel good or bad, or even more specific: what does hunger or pain feel like for another organism? Other species can differ widely from humans and from each other in terms of sensory, physical- and cognitive abilities. We can however likely infer at least certain aspects about some of these states, for example that pain involves a negative experience.

2.3.3 The hard question of consciousness

Subjective experience likely involves some kind of experiential center, or mind at a psychological level, and this is commonly studied through cognitive sciences. The physical brain is studied with neuroscience. The mind and the brain are often treated as two parallel causal lines of study (Fuster, 2003). It is challenging to see how these two can be joined together; today the only logical relationship between them is correlational. The hard question of consciousness (Chalmers, 1996) is to explain how and why the physical brain generates conscious experience, rather than just behavior or function. Even though science can contribute with empirical findings and correlates to consciousness, the solution to the hard problem will be a problem in causal explanation.

The solution to the hard problem can be either physicalist or non-physicalist. According to physicalism everything, including consciousness, is physical or supervenes on the physical. This means that all the different features that can be observed in the world, such as the psychological, biological or social, supervene on the physical: these features, or global properties, are such that they cannot differ without differing in their physical properties (Lewis, 1986). A physicalist solution hence implies that conscious experience can be explained in terms of physical processes. Assuming that physical processes by default are nonconscious, and a behavior can be explained in terms of physical processes then these processes should be able to do the causal explanatory work, and this renders the conscious experience causally superfluous. A non-physicalist solution, or dualism, implies that consciousness is explained in terms of something nonphysical. A reductive explanation can explain consciousness in terms of physical processes without having to refer to consciousness (Chalmers, 2010). A non-reductive explanation assumes that consciousness is a basic part of the explanation. A physicalist framework allows for consciousness to be an emergent phenomenon, but it still has to be able to be reduced to physical processes.

If we knew the answer to the hard question, then we would know the necessary and sufficient conditions for consciousness, and then we could answer the distribution question. However, it may be argued that it is not necessary to solve this very hard problem in order to be able to say something about the distribution problem (Dawkins, 2008). We can still find reliable correlates.

Many current scientific theories of consciousness assume that consciousness as a phenomenon can be explained in terms of physical processes, and that it can be reduced to physical principles. In the next section, I will present some examples of how consciousness may be physically realized.

2.4 Physical realizers of consciousness

Within many physicalist theories of mind it is common to assume that consciousness has something to do with activity in circuits of neurons in the brain. This belief is largely based on observations that conscious experience can be altered by factors that affect neurons such as brain lesions, brain abnormalities, and physical- and chemical brain stimulation. Because neurons and neural nets, or circuits, are often considered essential for human consciousness it is relevant to take a closer look at what these claims amount to and how they relate to animal consciousness.

2.4.1 Neurons and neural nets

Living organisms consist of many different kinds of cells. Neurons differ from other body cells in terms of cellular structure, function and replication. Neurons can receive, process and transmit signals. They receive signals from other neurons through structures called dendrites. The cell body computes these signals and if it reaches a threshold value then the cell fires an axon potential, an electrical signal that travels along the axon to other neurons. This allows them to communicate across vast distances in the body. At the end of the axon the cell releases chemical molecules into the synaptic cleft, the synapse, which reaches the dendrites of other neurons. Hence neurons use electro-chemical processes. Other kinds of body cells communicate only through chemicals and therefore only reach cells in close proximity. Furthermore, other body cells are independent units that can do their job alone, whereas neurons have a more global function in the sense that they perform their jobs through sending signals to other neurons.

There are different kinds of neurons. Sensory neurons receive external inputs that they transmit to the brain and motor neurons receive inputs from the brain and cause outputs in the body. Interneurons are neurons that connect neurons to other neurons and hence constitute the neural computational levels in the networks of neurons. Neurons are found in the brain, the spinal cord and the peripheral nervous system, but certain kinds are also found inside internal organs in the body, such as the heart or the gut. Many neurons are involved in what is considered to be automatic processes, hence not all neurons are assumed to play a role for consciousness, at least not all the time.

The neurons together compose networks. The networks of neurons in the human brain is vast and consists of 86 billion neurons (Azevedo et al., 2009) and each connects with up to 30 000 other neurons. These connections are not random, but have developed gradually and are constantly being modified. At the network level signals travel between neurons. A signal from a neuron will either inhibit or excite the next neuron, determining the path of the signal. Neurons exhibit advanced computational abilities and it is believed that they use their firing rates and time of spike to encode information.

Many theories of consciousness assume that neurons or networks of neurons play a role for conscious experience, but make different assumptions regarding their role.

2.4.2 Computational aspects

The computational aspect of neurons is essential in many theories. Computation is any kind of calculation or information processing. It can be argued that neural nets are computational systems and that the right kind of computational structure is sufficient in order to have a mind and mental properties (Chalmers, 2012). Computation can thus provide a general framework for explaining cognitive processes and behavior. However, if the brain is to be computational is has to be possible to describe how the computation can be physically realized; the causal structure of the system needs to be reflected in the formal structure of the computation. It is not enough to understand the abstract mathematical theory of computation.

Different theories make different assumptions about how consciousness is related to neuronal computations: as an intrinsic property of computation, a consequence of the computation, an emergent property at a higher neurophysiological level or if it has to do with the data that are computed over. Not all computations are assumed to be involved in conscious experience, and different theories have different requirements as to what makes certain computations conscious. For example, some require that the computations occur in specific brain areas (subcortical areas: Panksepp, 2011; neocortex: LeDoux, 2015), that the quantity of computations is sufficiently high or that the connections are sufficiently complex such that the brain can learn to become conscious through meta-representations (Cleeremans, 2011) or that the signals are spread in a specific way or to a sufficiently large area (Baars, 1997). Consciousness can also be related to aspects of the information itself generated by the computations such as the amount of integrated information and the informational relationships it generates (Balduzzi & Tononi, 2009). Recent studies indicate that the computational capabilities of neurons are more advanced and more dynamic than previously believed (Sardi et. al, 2017) and hence there may be aspects that have not yet been explored.

Computations alone may not be sufficient to generate conscious experience; the biological substrate itself may play an essential role. Furthermore, the mind may be more embodied than what is often assumed hence it is not only the brain that is involved but also peripheral parts of the nervous system (Nöe, 2009). There is today no agreement on brain mechanisms of consciousness (Merker, 2017). Consciousness may involve different parts and mechanisms than what is presented here. There are also other views on consciousness. Panpsychism, the idea that consciousness is an intrinsic feature of all things, like for example mass, is getting increased attention again (Strawson, 2006). Another position, which challenges the idea that consciousness is computation, is that consciousness is derived from wave function collapse in microtubules (Hameroff & Penrose, 2014). These different ideas need not be mutually exclusive.

The involvement of a nervous system at some level still seems a strong hypothesis and this warrants a closer like at the evolutionary history of the nervous system.

2.4.3 The origin of neural systems

Various single-celled eukaryotes can perform sophisticated behaviors such as phototaxis, gravitaxis and chemotaxis, without having neurons. In these organisms sensory inputs directly affect the motor behavior of the cell, such as the activity of cilia (Jekely, 2011). The efficiency of sensory-to-motor transformation in these organisms is low. The evolution of neurons increased this efficiency by allowing signal amplification and fast-long-range communication between sensory and motor cells. Some organisms found today have retained low levels of integration, such as cnidarians (phylum of over 10 000 marine species) and the ciliated larvae of some marine invertebrates. Other organisms have developed highly integrated systems.

Building and maintaining nervous systems have extremely high costs (Chittka & Niven, 2009), and this strongly shapes neuroanatomy. Overcapacity is severely penalized. Many animals do not have a cortex. Different species face different challenges and thus need to solve different kinds of problems. One can argue that they may have developed structures in the brain that perform analogue functions, such as in insects (Barron & Klein, 2015). The number of neurons in species differs vastly from a few hundred in nematodes to hundreds of billions in the elephant (humans: 86 billion). But, having a nervous system may not necessarily entail having conscious experiences.

2.5 Methods to study the presence of consciousness

An answer to the distribution question requires criteria for the presence of consciousness. Only the subject herself has direct access to and certain evidence of her own experiences. The subjective nature of consciousness has led many to believe that there can be no objective science of consciousness (Nagel, 1974). But science can, however, tell us about many important features, such as a bat's senses (Akins, 1993). Searle (2000) argues that there are two different senses of the subjective/objective distinction: an epistemic- and an ontological sense. We can have an epistemically objective science about ontological subjectivity such as consciousness.

Consciousness is studied through three correlates: verbal report, neural correlates of consciousness (NCC) and behavior. These can be considered either minimal requirements or as positive markers of consciousness.

2.5.1 Verbal report

A verbal report is based on the assumption that there is a correlative relationship between conscious states and words uttered. Between humans this relationship is often taken to be causal, and treated as direct evidence of conscious experience. For example, when a subject reports: "Ouch, this hurts" this means that the subject consciously feels a pain. It is however, an indirect measure and therefore we cannot know for sure. The subject could be lying, she could misjudge her experience due to memory failure or fail to express it correctly. A subject can even wrongly infer from her behavior that she was conscious. It is questionable whether or not a subject can always know if and of what she was conscious (Newell & Shanks, 2014). It takes time to generate a verbal report hence the words almost always follow after the experience. Explicit memory therefore plays an important role. Humans can report having conscious experiences during dreaming in sleep when woken up (Siclari et al., 2017), but most of the time we forget about these and may mistakenly deny having them.

Verbal report of conscious experience requires referential communication. Human language contains symbols, concepts and labels of objects, events and experiences, which allows for verbalization and communication independent of time and place. For example, a label such as "hunger" captures something that is difficult to describe and communicate in terms of the underlying phenomenological experience, both sensations and motivations.

Ned Block (2007; 2011) distinguishes between access consciousness and phenomenological consciousness. He claims that it is likely that there are some phenomenological experiences that we cannot access with words and verbal thoughts, and therefore these are not available for verbal report – not even in the sense that you can identify whether or not you had them. Based on his theory a human can have phenomenal experiences that in a sense are not available to them. Dennett (1995) however objects to this view, claiming that access is just what consciousness is about and that what is involved is rather a difference in richness of content and degree of influence.

Non-human animals do not possess human language and are therefore not able to generate verbal reports of the human kind. We do not even know if animals use symbols or if this makes humans unique (Deacon, 1997). Symbols may have originated as the names of perceptual categories that were based on iconic and categorical representations (Harnad, 1987). There is evidence of animal species that have the ability to represent the experienced world at higher levels of abstraction (jays: Clayton et al, 2006; bees: Gallistel, 2008) hence these representational abilities may be evolutionary ancient (Gallistel, 2011). The use of symbols may be a capacity for linking sounds or gestures arbitrary to specific concepts (Christiansen and Kirby, 2003).

It is however, very difficult to determine whether or not something is a manifestation of symbolic processes, whether in early human archeological artifacts (Balter, 2009) or in other species, such as in birdsong (Berwick et al., 2011) or the different calls made by prairie dogs to alert for different predators. Chimps have successfully learned to communicate with humans with many signs in American Sign Language (Gardner & Gardner, 1969). This involved names of objects from imitation, but even the invention of strings of two or more signs, in addition to spontaneous naming and transfer to new referents. A recent study of gestures in Chimpanzees and Bonobos found that they share not only the physical form of many of the gestures, but also many gesture meanings (Graham et al., 2018). The Grey Parrot Alex was able to label objects, properties of objects, even quantity, and combine these in communication with humans (Pepperberg, 2012). Fish and eels seemingly go hunting together and this appears to rely on communication about the hiding place of non-present prey (Bshary et al., 2006). It is however still considered a challenge to demonstrate that an animal is behaving based on subjective experiences rather than exhibiting learned behaviors.

It is not long ago that the conscious experience of infant humans was questioned due to their undeveloped cognitive capacities and lack of language. Today they are considered conscious (Fitzgerald, 2012), in the same way as patients suffering from aphasia or humans with language disabilities that prevent them from ever learning any language. There are few studies on deaf children that are not taught any language, but individual stories exist that seem to indicate that language is integral for explicit memory, abstract thinking and self-awareness (Schaller, 1991). It has also been suggested that storing memories, developing language and the construction of thoughts can be realized through sensory modalities, even several modalities, for example in people with autism (Grandin, 2009). Also in animals, it is argued that thinking can take place without words (Bermudez, 2003), but without a common language this can only be manifested through behavior.

2.5.2 Neural correlates

Consciousness is also studied through observing what is referred to as "neural correlates of consciousness (NCCs)." The quest for NCCs involves identifying neural structures and mechanisms, and the neural events or processes, which are assumed to be involved during conscious experience. These NCCs are correlated with human verbal report, which itself is a correlate. Typically, researchers look at what is physically going on in the brain at the time the subject reports having conscious experiences. It is then assumed that these human NCCs are necessary for the subject to have a conscious experience. Many such NCCs have been proposed over the years. Some are characterized by the kind of activity or activity in specific neurons or brain areas (see Chalmers, 2010, p. 60 for a list). The activity of brain areas or circuits can be investigated by the use of different imaging techniques, such as scalp electroencephalogram (EEG) or functional imaging such as blood oxygen level dependent (BOLD) functional magnetic resonance imaging (fMRI) (Reese et al., 2002). It is however, not always clear what kind of information we can actually extract from these images and how to interpret them (Shulman, 2013).

The search for evidence of consciousness in animals is based on looking for similar NCCs in them. The necessary and sufficient conditions for human consciousness still however remain unknown and therefore the common NCCs for humans are only assumptions. And even if these *are* related to consciousness in humans, we cannot know if they are related to consciousness in animals because we do not have a verbal report from the animals to match it with. Furthermore, there are no criteria as to necessary degree of similarity. Using human NCCs as criteria for animal consciousness excludes the possibility that other species may have conscious experiences that are realized by different NCCs; as in the case of parallel evolution (Bronfman et al., 2016; Godfrey-Smith, 2016). If consciousness is a gradual phenomenon rather than "all-or-nothing" using human criteria for consciousness might also rule out species that only have consciousness in a minimal sense or rudimentary experience (Vandekerckhove & Panksepp, 2011).

Animals cannot deliver a verbal report, but they can perform behavior. Studies with Deep Brain Stimulation (DBS) on animals have identified subcortical brain areas that exhibit punishing and rewarding properties, which affect behavior (Panksepp, 2015). Animals are willing to work to obtain this stimulation. It fact, this has been demonstrated in all the vertebrates studied. Panksepp (2015) argues that this is the most promising entry-point into the neural understanding of affective (conscious) experience.

2.5.3 Behavior

2.5.3.1 Consciousness and behavior

The belief that conscious states have a causal effect on behavior is part of common folk psychology. But behavior is a correlate of conscious experience and can only provide indirect evidence. Behavior encompasses anything an animal does such as bodily movement and posture, facial expression and vocalizations. On this definition a verbal report can be considered a special case of behavior. Inner physiological measurements such as stress hormones, heart rate and core temperature may also play a role for conscious experience, either as cause or effect and hence play an indirect role for behavior. For example, if a person puts her hand on a hot plate the following observations can be made: The person quickly withdraws her hand and screams, and heart rate and stress hormone levels increase. We assume that these observations correlate with the subject feeling pain. We still however need to correlate the behavior with a verbal report.

If conscious states play a causal role for behavior then it should in principle have measurable effects on behavior (Allen & Trestman, 2014). And then it is possible to answer the distribution question with science. If on the other hand, consciousness is an epiphenomenon and does not have measurable effects on behavior, then it is difficult to see how consciousness could have been adaptive and selected through evolution (Allen & Bekoff, 1997). But even if consciousness is not a product of evolution, as for example in panpsychism, it could still be applied by evolution, for example like mass, and play a causal role for behavior.

It is however very challenging to provide evidence that a behavior is caused by a conscious process. For example chimps seem to be able to understand many aspects of other minds such as goals, intentions, perception and knowledge (see review by Call & Tomasello, 2008). Insects demonstrate complex cognitive abilities (Barron & Klein, 2015). Rats can be trained to respond to which of two tones is louder, by pressing a button (Akrami et al., 2018). The critical issue is whether or not the observed behaviors require conscious states or are merely examples of complex associate learning that does not require conscious states.

2.5.3.2 Pain and behavior

A feature of consciousness is its valence in terms of positive or negative experience. Pain is considered a negative experience that guides behavior. All kinds of organisms, from single-cells to mammals exhibit protective reactions against injury and damage, from simple withdrawal movements to complex protective behaviors. One of the earliest developments of nerve cells was the ability to sense tissue damage. Pain is commonly separated into two components: nociception and pain perception. Both components are part of the definition of pain by the International Association of the Study of Pain (IASP). Nociception refers to the nonconscious physiological processes that are involved in detection and response to noxious stimuli (Cervero, 2012). Pain perception refers to the conscious experience of pain. Nociceptive activity is present in all animals. Pain perception however, is under discussion. The question is whether or not stimulation of a sensory cell, or a system of neural cells, involves conscious experience. Broom (1998) argues that even nociceptive processes that do not require high-level processing in the brain involve subjective negative feelings.

Humans know from subjective experience that pain is aversive. We know pain alters our behavior and that we learn from painful experiences. We are also familiar with the typical human pain behaviors and physiological responses. Many animals respond similarly, both behaviorally and physiologically. Many studies address pain perception in animals and it is well documented that animals display facial, bodily, behavioral and physiological indicators of pain (castration in pigs: Weary et al, 1998; beak trimming and bone breaking in poultry: Gentle, 2011; tail docking in pigs, sheep and cattle: Sutherland & Tucker, 2011; hot iron branding in horses: Erber et al., 2012; exposure to various kinds of noxious stimuli in fish: Malafoglia et al, 2013; facial indicators in mice: Miller et al., 2016). Pain has been shown to alter their behavior and the administration of pain relievers reverses these effects (Andrews et al., 2012; Allen et al., 2013). Fish have been shown to respond to analgesics (Sneddon et al., 2003). Furthermore, animals will trade-off food in order to avoid painful areas or obtain painkillers (Millsopp & Laming, 2008). They will self-administer opiates when in pain, due to the pain itself, and not the rewarding effect of the opiates (Colpaert et al., 2001). Pain perception in fish is particularly under debate. Whether or not nociceptive processes involve pain is a central, but there is growing scientific evidence that fish can perceive pain and experience at least some of the negative aspects (Sneddon, 2011). However, Rose et al. (2012) disagree with this conclusion. The assumption that at least many animals consciously experience pain has great explanatory power in predicting their behavior.

When we study pain in animals we often start by situations we can relate to. Different species however, vary in terms of physiology and behavioral responses. Which stimuli elicit pain and the tolerance of stimuli may differ greatly from humans. Furthermore, many animals go through great efforts to hide pain in order to avoid signaling weakness to conspecifics or predators. Humans on the other hand can benefit from displaying pain to communicate it to others.

2.5.3.3 Morgan's Canon

Behavior is in an indirect measure. A skeptic can always claim that if behavior can be explained in terms of nonconscious process then conscious states should not be posited (LeDoux, 2015). This view often refers to Morgan's canon. This states the following: "In no case is an animal activity to be interpreted in terms of higher psychological processes, if it can be fairly interpreted in terms of processes which stand lower in the scale of psychological evolution and development" (Morgan, 1903, p. 59). The canon can be seen as C. Lloyd Morgan's adaptation of the principle of Occam's razor and Hamilton's law, specifically for comparative psychology (Karin-D'Arcy, 2005). Occam's razor and Hamilton's law state that one should not postulate theoretical entities that need not exist; one should aim for simpler explanations.

Morgan's canon was intended for the study of animal psychological functions. The idea was that when comparing these to human psychological functions, one should not postulate higher functions, or higher faculties, if not necessary for the explanation. The statement has however, been interpreted in different ways. It has been taken to mean that "lower in the scale of psychological evolution" refers to states that have no psychology at all in the sense of no conscious experience (LeDoux, 2015). On this view, taken together with Occam's razor and Hamilton's law of parsimony (1856, see Karen-D'Arcy, 2005 for discussion), conscious states are taken to be examples of theoretical entities that should not be postulated. But it is not clear that the intent of the canon was to distinguish at an ontological level, i.e. between conscious and nonconscious states. Rather, it can be argued that the canon was intended for distinguishing between different kinds of psychological states, such as whether or not an animal is capable of intentionally deceiving other animals. Interpreted charitably, what the canon is asking is to avoid our tendency to project from ourselves and see more complexity than is actually there.

Even as a tool for comparative psychology, the canon is criticized and the need for such principles at all has been questioned (Mercado, 2016). According to Starzak (2016), there is no reason to assume that an explanation on a scale of psychological evolution and development should be better at some point than another. An explanation should be preferred based on its explanatory power and how likely it is to be true. Simplicity is considered a virtue in science, but what counts as simplicity must be clear of underlying assumptions about human uniqueness. There will always be cases where simpler explanations are not necessarily more correct that complex ones. Given the nature of the brain, it may well be that the complex explanation is the right one (Broom, 2010). From an evolutionary point of view, there is biological continuity between species. Hence the idea that consciousness has evolved just like other processes is scientifically more plausible than the idea that humans differ essentially from all other animals, even our closest relatives the great apes. Using Occam's razor here does not simplify ontologically, rather it leaves us with two things to explain instead of one.

2.6 Current beliefs about animal consciousness

In the Western scientific community today the generally accepted idea is that neural systems are essential for conscious experience. In 2012, a prominent group of cognitive neuroscientist, neuropharmacologists neurophysiologists, neuroanatomists and computational neuroscientists signed "The Cambridge Declaration on Consciousness" (Low et al., 2012). It was publicly proclaimed in the presence of Stephen Hawking. The declaration states that:

> The absence of a neocortex does not appear to preclude an organism from experiencing affective states. Convergent evidence indicates that non-human animals have the neuroanatomical, neurochemical, and neurophysiological structures of conscious states along with the capacity to exhibit intentional behaviors. Consequently, the weight of the evidence indicates that humans are not unique in possessing the neurological substrates that generate consciousness. Non-human animals, including all mammals and birds, and many other creatures, including octopuses, also possess these neurological substrates.

Conscious experience in fish, invertebrates and insects is still actively debated (vertebrates: Sneddon, 2004, Elwood, 2011; fish: Sneddon, 2011,Rose, 2012; insects: Barron & Klein, 2016). It does however seem that as research on their physiology, cognition and behavior progresses, there is gradual acceptance that these animals are conscious. This tendency was confirmed at the recent multidisciplinary conference on "Animal Consciousness" at NYU in November 2017.

There are however objections from camps within the scientific community to the idea that animals can have conscious experiences, and that claim that this view is not based on science (LeDoux, 2015). These objections hold a firm grip within many communities involved in the instrumental use of animals such as food production and research. This prevents progress. The debate on animal consciousness is complex in the sense that it involves the use of different definitions, terminology, assumptions and theories, and kinds of questions being asked. It is important to get an overview and disentangle the various factors involved in order to be able to identify tensions, clarify positions and their relation to science. These arguments need to be addressed and responded to, to resolve what can settle the debate.

A highly prominent proponent of these objections is the Canadian neuroscientist Joseph LeDoux. LeDoux has been the most funded researcher on emotion in North America and for decades he has performed extensive research on animals related to behavior, learning, memory and emotion, specifically focusing on processes related the feelings of fear and anxiety. Recently he has been very explicit in expressing that what he previously referred to as "fear" in his models of animals do not refer to the *feeling* of fear known to humans, but rather to automatic nonconscious processes in the animal that detect and respond to threats and danger (LeDoux, 2012; LeDoux, 2014). His claim is that he can explain animal behavior in terms of neuroscience. Therefore, based on scientific evidence, conscious states should not be posited in animals to explain their behavior. He goes even further by proposing a model for consciousness, which excludes animals altogether.

LeDoux is a highly respected scientist, with deep knowledge within both neuroscience and cognitive science. His work and his positions are likely to have significant influence on the attitudes of many people. LeDoux is willing to address and discuss difficult issues that many others evade. Even though his argument against animal consciousness is classic in form, there is no clear procedure for how to respond to it. And this is the argument that all species under consideration are brought up against. He is very explicit in his allegation that animal consciousness is being declared without rigorous and compelling scientific evidence, whereas his model and his claims are based on science. This is important to respond to. There are probably many reasons that more species are gradually being considered conscious, one being more knowledge about them. But there has also been a general shift in society towards accepting other beings as equal, such as women and children. Even though discussions like the one raised by LeDoux will play a role in this gradual shift, his claims are important to take seriously because we clearly want the shift to be based on science and because his views retard an already fuzzy debate. His points raise some very important and interesting issues that can help tidy the debate and identify how inquiries on consciousness and animals should best proceed. In this way his contribution can improve the debate.

2.7 Introduction to LeDoux's book

LeDoux has written a book titled "Anxious. Using the Brain to Understand and Treat Fear and Anxiety"(2015). This book is a contribution to the understanding and treatment of fear and anxiety in humans. It is a thorough review of neuronal processes, neural circuits and systems in the brain that are involved when an organism is exposed to threat and danger. He argues that both humans and animals can detect and respond nonconsciously to danger. But only in humans do these processes contribute to a system of circuits and cognitive functions that allow for conscious experience of these nonconscious processes. Conscious experience functions as an interpreter, a storyteller.

Animals are not conscious, and therefore animal models allow insight into only some of the processes that influence conscious experiences of fear and anxiety in humans. "Fear" is an ambiguous term in the literature, sometimes referring to a conscious experience and sometimes to nonconscious processes. LeDoux believes that this leads to conflation of systems that detect and respond to danger with systems that give rise to subjective conscious feelings of fear (LeDoux, 2014). This calls for a "sharper conceptualization of what is being studied" (LeDoux, 2015, p.37), and this is what he attempts to do with this book.

LeDoux claims that science cannot tell us about consciousness in other species. He also claims that he can explain animal behavior in terms of neuroscience. He presents his own theory of consciousness, which excludes animals. Both of his claims have consequences for the distribution question. The first one regards what science can tell us about the presence of consciousness in other minds, of humans and other species. LeDoux obviously believes that science can tell us something about consciousness; after all he has a scientific theory of consciousness. But this is only applicable to humans because he knows for sure that humans are conscious, whereas no empirical fact can tell him that animals are conscious. I will examine his positions, their consistency and the consequences for the distribution question.

LeDoux's second claim touches on several issues. One regards explanations of phenomena in general; the factors involved and what kind of conclusions they warrant. The other regards his explanation of animal behavior and whether or not this is a plausible account of behavior. I argue that his explanation is not exhaustive. I then discuss his model of the relation between stimuli, consciousness and behavior. I identify two underlying assumptions, object to these and argue for an alternative model where consciousness is related to reinforcement and learning. This introduces a new marker of consciousness.

In the next section I will present LeDoux's first claim. Discussion will follow in the following section.

3 Presentation Claim 1

LeDoux believes that questions regarding conscious experience in animals cannot be answered by science (LeDoux, 2012; LeDoux, 2015). Conscious experience is subjective: only the subject has direct access and direct evidence. We will never have direct evidence of consciousness in other species, and therefore we can never know. Recently at a conference on Animal Consciousness at NYU (2017), when asked about whether or not he believes animals are conscious he answered the following: "As a person I do. As a scientist I don't." He believes the view that animals are conscious is an ethical position, not a scientific one (LeDoux, 2015, p. 48). The case for humans is different. He can know that other humans are conscious based on his own subjective experience and indirect evidence from science. There is hence an epistemic asymmetry between the case for humans and the case for animals. In order to understand his position it is necessary to look at his definition of consciousness, his requirements for consciousness and requirements for evidence of consciousness.

3.1 LeDoux and consciousness

3.1.1 What is a conscious state?

For a state to qualify as conscious in LeDoux's theory, it is not sufficient that the organism is awake, alert and responsive; the organism also has to be explicitly aware of what it is experiencing and that it itself is having the experience. He requires the following (LeDoux, 2015, p. 230):

> This can only happen in organisms that have the capacity to both be aware of brain representations of internal and external events and to know in a personal, autobiographical sense that the event is happening to them – someone has to be home in the brain in order to feel fear when the defensive circuit knocks on the door.

An explicit concept of self is essential. There is a difference between being aware of the presence of a stimulus and being aware that it is yourself who is having the experience: "To experience fear is to know that YOU are in danger" (LeDoux, 2015, p. 50).

LeDoux does not reserve the term "mental" for conscious experiences. Mental, in his terminology, refers to the use of internal representations of the world hence nonconscious does not mean nonmental in his view. The brain performs complex cognitive mental operations to control behavior, but this is different from having the ability to have conscious experiences. When a mental state is conscious, he refers to this as "mental state consciousness".

3.1.2 The making of conscious states

LeDoux presents a positive theory of consciousness, which includes a list of parts and events that occur in an organism when it encounters a dangerous stimulus (LeDoux, 2015, p. 227). Each event requires that the previous events on the list are present:

1. Representation of the stimulus in the brain – nonconscious
2. Defensive circuit activation. This causes physiological changes and behavior - nonconscious a-noetic state
3. Attention/Working memory. This allows for conscious knowledge of the stimulus - conscious perception
4. Semantic memory. This allows for factual consciousness: object recognition (this is a snake), knowledge about the thing (snakes are dangerous) - noetic consciousness
5. Episodic memory. This allows for the 'when where and what and YOU (self)' in the episode – still not a conscious state of fear
6. Monitor and recognize that these ingredients are indicators of fear → categorize and label fear – autonoetic conscious state of fear

LeDoux distinguishes between different kinds of knowledge-states. Noetic state of fear (No. 1-4) is in principle possible in nonhuman animals, but it requires that the brain is conscious of its own activities, and this we can never know due to lack of direct evidence. His explanation of animal behavior, and also quite a lot of human behavior, involves only No. 1 and 2. A conscious mental state is first present in No. 6. Autonoetic consciousness, which most likely is exclusive to humans, requires that the brain can apprehend that the event is happening to itself (LeDoux, 2015, p. 46):

> Thus, when a defensive survival circuit has been activated in your brain and its consequences linked to the present stimulus and to your memories regarding it and similar stimuli, all in relation to your awareness that the event is happening to YOU, a feeling of fear arises.

According to his theory, consciousness emerges from non-conscious parts. He illustrates this with an analogy of soup: the flavor of a soup emerges from ingredients that are themselves "non-soup". The flavor can change in quantity (or intensity) from adding salt or pepper, but it can also change qualitatively form changing the ingredients, just like the quality of an emotion can change based on its ingredients. The necessary ingredients for consciousness are: sensory processing, survival circuit activity, brain arousal, body response feedback, semantic, episodic, autobiographic and implicit memory, executive function such as attention, monitoring, labeling and attributing. The cooking pot is working memory, and the soup itself is the conscious feeling. Feelings like fear and anxiety are a consequence of the cognitive processing of non-emotional ingredients. Consciousness functions as an interpreter of nonconscious processes, and figures as a storyteller.

Language plays an important, perhaps essential role, for at least the last two requirements. He believes that a feeling such as fear requires the concept of fear, and this is based on words and the meaning of these words in our mind. Language allows for labeling of categories such as experiences, and may also be necessary to allow for the concept of self. He refers to Dennett (1991) who says that language is the tracks on which thoughts can travel. In humans at least, language has likely allowed consciousness to emerge in this way.

These cognitive functions require specific neurophysiological structures and functions, such as the ability for self-representation. The neocortex plays an essential role and the necessary parts are exclusive to humans. In fact he considers these areas so essential for conscious experience that before these mature in the human brain sufficiently to realize cognition and language, there may not be conscious emotion. The defensive circuits in subcortical brain areas mature earlier than these cortical circuits and therefore human infants may behave emotionally before they can actually feel emotion.

LeDoux discusses other views on what constitutes a conscious state. Panksepp (2011) for example, argues that primitive conscious states can be anoetic, i.e. without explicit awareness. These are realized by subcortical structures. Inherited basic affect programs are realized by the subcortical limbic system and arousal systems. LeDoux however argues that these affect programs are hypothetical structures; placeholders for brain mechanisms. More likely, according to LeDoux, emotions are psychologically, cognitively constructed concepts that are made possible by language and beliefs.

3.2 LeDoux on the epistemology of consciousness

LeDoux assigns consciousness in the world starting with himself. He then works outward from this. He knows that he himself is conscious, based on direct access to his own experiences. This direct evidence, together with indirect evidence of other humans, gives him sufficient evidence that other humans are conscious as well. This is based on two criteria. The first is physiological similarities to himself: "If my brain can be conscious so can yours" (LeDoux, 2015, p. 49). The second criteria is verbal report: "..verbal self-reports are the best way to verify and compare conscious experiences between two organisms" (LeDoux, 2015, p. 49). A verbal report is not always perfect, but it is the best there is.

3.2.1 Correlates of consciousness

LeDoux is conscious and because other humans have very similar physiology to him, and can provide a verbal report, he knows they are conscious too. But animals are not similar enough. In fact they differ significantly: "The human brain differs in significant ways from even our closest primate cousins, not so much in terms of the areas that are present but in their patterns of connectivity and cellular organization" (LeDoux, 2015, p. 49). Therefore their physiology cannot tell us about consciousness.

Humans can provide a verbal report. Animals do not possess human language. They would have to demonstrate by a behavioral substitute. But LeDoux requires that they demonstrate that their behavior is caused by consciousness according to his definition. Hence it is not sufficient that an animal non-verbally express that they are aware of an object; they have to be able to show that they are aware that they themselves are having this experience. Evidence of this is not possible to find in behavioral data according to LeDoux. You can never know if it is *this* awareness that plays a causal role for behavior. This is why he is not willing to accept the behavioral evidence presented by Panksepp (2015), where animals are shown to behave to obtain stimulation of reward sites. We cannot know by direct evidence if the observed behavior is caused by conscious experience with explicit content, or by nonconscious processes.

LeDoux has a model (LeDoux, 2015, p. 45) of behavior and consciousness. In this model the nonconscious processes that underlie behavior also play a role for generating conscious experiences. This makes it impossible to know if it is the underlying nonconscious processes or conscious processes that cause behavior. The only way to determine which processes are conscious is by asking the subject. Animal behavior can therefore never be evidence of conscious experience. It can only show us that the brain registered the stimulus in a meaningful way. LeDoux goes even further and claims that because these underlying processes can account for behavior without referral to conscious processes, there is no need for consciousness to explain animal behavior. Animals can perform all the necessary behavior they need in life nonconsciously (LeDoux, 2015, p. 47): "… animals can satisfy nutritional and fluid needs by consuming food and drink, have sexual intercourse, writhe when injured, and freeze or flee when threatened, all without the necessity of conscious awareness that they are doing so." They rely on innate responses, conditioned reactions and nonconscious cognitive processes. Hence these behaviors are not evidence of conscious experience.

Nonverbal responses, such as physiological reactions and behavior, do not work for humans either, argues LeDoux. They cannot on their own reveal the difference between conscious and nonconscious, it has to be supplied by a verbal report. Humans can detect and respond to threat nonconsciously (Whalen et al., 1998). Threatening stimuli have the advantage that they elicit automatic bodily responses such as changes in blood pressure, heart rate, respiration and perspiration, and they can do so without the subject being aware of the stimulus. Hence these are reliable indicators that the stimuli have been meaningfully processed. Stimuli not consciously perceived can also affect behavior, such as in cases of blindsight (Stoerig & Cowey, 2007).

Also the human verbal report has challenges. Humans cannot rely on their intuitions to determine the role of conscious experience on their behavior because we often make mistakes. For example, there are many kinds of things we learn and do which involve implicit, nonconscious processes such as syntactic parsing of sentences, depth perception and instrumentally reinforced behavior, even though we feel as if our conscious experience plays a role. Similarly, in situations that require fast action, such as jumping away from a snake, we often attribute conscious experience, such as fear, to cause behavior even though the behavior was executed by nonconscious processes before conscious experience had time to unfold and play a causal role.

Not only are we prone to misattribute our own consciousness, but we also do this when we interpret other animals. Many species act in ways that are very similar to humans and we wrongly interpret that conscious experience plays a causal role for their behavior. This is also the case with signs of intelligence and cognition. This is argument by analogy, not scientific evidence. Humans are social animals wired to interpret other humans in order to understand and predict their behavior. We are thus prone to anthropomorphize, i.e. interpret behavior with regards to human intentionality and motivation. The folk-psychological view is that there is a relation between our conscious experience and our behavior, and this can automatically transfer to non-human animals, even to nonliving objects that seemingly move with intentional movement. For example, figures on a computer screen can be described as “aggressive triangles” and “fearful circles” (Heider & Simmel, 1944).

LeDoux also makes some speculations about consciousness in animals: “Even if consciousness is present in other animals in some form, it can’t exists in the way made possible in the human brain” (LeDoux, 2015, p. 50-51). Inspired by Morgan’s Canon, he also presents a requirement for behavior to indicate consciousness (LeDoux, 2015, p. 47):

> Consciousness should therefore only be attributed to an organism if there is both compelling evidence that the behavior expressed by the organism depends on consciousness and compelling evidence that the behavior cannot be explained in terms of nonconscious processes.

He also makes some speculations about the phenomenal content of conscious experience, for example: “They (the animals) may experience something, but it is incorrect to assume that their experience is the same as, or even similar to, what a human experiences when his or her defensive survival circuit is active” (LeDoux, 2015, p. 50). In a lecture he stated the following about the rats and the floor-shock they were subjected to in one of his experiments: “..the shock is relatively mild…it is not hugely painful to them…we put our hands on the grid all the time to verify that it is working properly" (LeDoux, 2011).

3.3 Conclusion

LeDoux's definition of consciousness requires that the subject have explicit knowledge about the experience and that it is herself who is having the experience. In humans, language allows for the necessary concept of self and labels of experiences. This requires cognitive resources that in turn are realized by specific neural structures, which can self-represent. He does not believe that animals have the necessary ingredients for conscious experience.

LeDoux has different requirements for evidence of consciousness in other humans and other species. For other humans, he relies on their similar physiology to himself and the verbal report. Animals however, differ too much in physiology. Furthermore, they cannot provide verbal or behavioral evidence that demonstrate consciousness in accordance with LeDoux's definition. Behavior is not a good correlate of consciousness. Humans often mistakenly attribute conscious experience to play a role for behavior both in themselves and in the interpretation of other animals. Furthermore, in his model, nonconscious processes contribute to conscious processes. Therefore, one can never know which is doing the explanatory work.

I will now analyze and discuss LeDoux's statements and their consistency.

4 Discussion Claim 1

LeDoux makes the following statements:

(1) In a conscious state the subject has explicit knowledge about both the experience and that the experience is happening to itself.

(2) Conscious experience requires specific ingredients in terms of neural structures and cognitive functions that can realize this explicit experience.

(3) LeDoux is conscious – by direct evidence.

(4) Scientific observations - by indirect evidence - can tell us about consciousness in other humans:

- Verbal report: they can tell us
- Neurophysiology: sufficiently similar neurophysiology to himself
- Behavior: difficult to demonstrate consciousness

(5) Scientific observations – by indirect evidence - cannot tell us about consciousness in other species:

- Verbal report: not available
- Physiology: too different from him
- Behavior: cannot demonstrate consciousness

First I will discuss the content of these claims. Then I will examine how they relate.

4.1 The content of his claims

4.1.1 (1) LeDoux's definition of consciousness

LeDoux believes that conscious experience is a psychological construction unique for the human mind. The human brain constructs a self that is the center for experiences, and functions as a storyteller. He distinguishes between the physiological nonconscious processes that contribute to the conscious state of fear, and the conscious state of fear, which is to be aware of these nonconscious processes. Conscious experience requires a self to experience these nonconscious processes *and* to have knowledge that it is happening to itself. Self-consciousness thus seems to be necessary for this view. This is a version of higher-order theories of consciousness (LeDoux & Brown, 2017).

There are two important issues to address. First, is it plausible to assume than an experiential center requires an explicit concept of self? If so, is language the only form of labeling or can there be some kind of non-linguistic self-reference? As mentioned in the introduction, it seems plausible that language developed from at least some kind of categorization system. If the first is not plausible, it is not necessary to examine the second one for my purposes. I will start by looking at human experiences and the role of language and labels.

4.1.1.1 *Labels, the self and the experiential center*

Language allows us to label experiences, categories and concepts. According to LeDoux a nonconscious process that is not labeled can affect behavior in both animals and humans. But a label for a nonconscious process allows the self to identify and recognize the nonconscious process, and through this process it emerges into consciousness. This explicit conscious self is the experiential center, and this must be in place for the self to have other conscious experiences. So the first label that must be in place is the label for the self.

A label for the self arguably relies on a concept of self. A concept can be a mental representation that plays a computational role through being embedded in systems of core cognition. LeDoux would not deny that these play such roles in animals. For example, bird navigation requires representations; there is nothing from what the sky looks like that has the content north (Carey, 2009). Concepts can however also be embedded in explicit knowledge systems, and these are likely the kinds of concepts LeDoux refers to. A concept as an abstract generalization must refer to certain instances of nonconscious processes in his model. This requires that first some nonconscious processes, probably self-referential, are categorized and then abstracted into a concept of self. This concept is then labeled and either in this process in itself, or as a consequence of this process, a conscious self is born.

In this account, the experiential center requires an explicit self. Two aspects of this center for subjective experience are the subjective and the phenomenal or qualitative (Godfrey-Smith, 2017). I will make the assumption that the explicit self in LeDoux's theory is both subjective and phenomenal. The explicit self is established and this constitutes the center for subjective experience. For conscious experiences to be present it contains both explicit knowledge about this self and the experience. The kind of knowledge necessary to constitute a conscious experience is a major area of tension between theories. LeDoux requires autonoetic knowledge, whereas Vandekerckhove & Panksepp (2011) argue for anoetic knowledge. I will come back to this later when I examine which kind of knowledge is necessary to explain behavior.

The idea that the experiential center relies on an explicit self relies on the premise that labels are constitutive for conscious experiences. I will make the assumption that the label for the self plays its role in a way that does not differ significantly from the other labels, other than the fact that it must come first. Are labels constitutive of human conscious experience?

Several objections can be raised. First of all, there are many examples of cases that would not qualify as conscious in LeDoux's theory. Humans arguably have conscious experiences that we sometimes are not able to label because we just can't figure out what we are feeling, or because language does not have adequate labels for it. We also have conscious experiences where labels are not consciously involved because they just don't come to mind – even though labels do exist and we do know about them. For example, there are cases where we feel things, and even though we know exactly *what* we are feeling we do not understand the feeling in a way that allows us to label it. We are not able to find the adequate label even though we try. Yet we can still feel them and think about them in terms of the experience itself. These experiences can be simple sensations or complex feelings, vague and fuzzy but also very clear, yet introspection does not reveal to us a label. LeDoux could argue that I am indeed applying a label termed "unknown feeling". However, his theory specifically states that the labeling process causes the conscious experience to emerge, and this arguably requires identification and recognition.

I believe I can distinguish between conscious experiences with and without labels. I can have sensory experiences that do not involve labels. Not because I do not know the labels, as in the previous examples, but because they are not on my mind. Only the taste of the cookie is on my mind, or the tune of the melody. I am not having explicit thoughts about the experience, even though I could. Through mindfulness and meditation techniques people claim to be able to have conscious experiences that do not involve a self and a storyteller (Langer, 1989). For example, one of the techniques is to be conscious of breathing activity only. There are also sensory experiences that language does not have labels for because they are impossible to describe. For example, the sensations I feel in my stomach when I am hungry. The label is a crude label for a category that also involves intentions and motivations, and which can be communicated to others. But labels cannot be used to describe the sensations themselves.

Most experiences seem to involve a mix of labeled and non-labeled experiences. For example, I can step on a tack and feel a sharp pain, and in the beginning all there is is a feeling of pain. I do not think about who and what, at least not that I am aware of. The pain is accessed directly; there are no labels and no thoughts. I do not have to know what a pain is – neither as a symbol or a concept – to know that it hurts. My experience does not involve explicit knowledge of the kind: "This is an experience of pain and I am having this experience." In fact, it is first when the pain subsides that my mind is able to entertain the experience in labels such as: "Ai this hurts", or "I must have stepped on a tack."

These experiences demonstrate that LeDoux's definition of consciousness is too narrow because it excludes many human experiences, at least many of mine. It may also exclude animal experiences. LeDoux's version of consciousness can be a part of a conscious experience, but is not necessary for a conscious experience.

Secondly, LeDoux's claims imply that we come to know our conscious experiences through our language. Language allows us to identify and recognize experiences. But this seems to contradict with how humans learn new labels. Adults teach children labels by putting words to what they are feeling. For example, if a child hurts himself then we typically say: "Ouch, that hurts, you are in pain." According to LeDoux these labels must accompany the nonconscious process involved in order to emerge into consciousness. This implies that the child cannot feel pain before she has learned to label it. Similarly, if I as an adult have never experienced being dizzy, and then I experience this for the first time. It seems odd that I cannot have this experience before someone has taught me the label for it.

Thirdly, a verbal report refers to past experiences, at least it is difficult to conceive of examples where words do not follow after the experience. It therefore involves mental time travel and hence explicit memory. Furthermore it involves language, which consists of labels. Thinking about an experience in terms of language necessarily presents the experience as an explicit experience. When I am asked about a conscious experience, I have to label the experiences in order to be able to verbalize it in thought and communication, and the answer will include a "me" along with other labels. But I can truly say that labels were not part of my immediate experience of the cookie or the melody, they are only part of my description of the experience.

LeDoux might argue that even though my intuitions tell me differently, the labels and thoughts were baked into my experience. But this then implies that these labels can play a role nonconsciously, because they were not part of my conscious experience to my knowledge. LeDoux cannot accept that because the whole labeling business *is* consciousness; I have to know about it. But then this means that either my conscious experiences involve labels without me knowing about it, or that I think I have certain conscious experiences which in fact are not conscious. Hence my ability for introspection is not very trustworthy and this is a challenge for the verbal report.

An example that may be a case of an experiential self without an explicit self is the deaf man that grew up without learning sign language (Schaller, 1991). At the age of 27, after much work with a teacher, he finally grasped the idea of symbols and learned sign language. He described the time before this as "darkness" and that "there was no I". Yet he had still experienced fear and pain, hence he was arguably not without an experiential center.

For LeDoux it seems that it is the thoughts that interpret the experience that constitutes the conscious experience. It is not until the experience is explicitly recalled that it becomes conscious, even though we feel as if we were conscious at the time. This seems to imply that we can become conscious of experiences at a much later time than they seemingly occurred. Here an example by Susan Blackmore (2016) comes to mind: She is climbing a mountain and automatically looks at her watch. A moment later she asks herself if she was conscious a moment ago, and it is then she recalls the memory and becomes aware of her behavior. Was her clock-watching a moment ago nonconscious and became conscious when she thought about it? The memory first comes into consciousness when she asks herself the question. Or did she have an experience back then, but that was immediately forgotten – replaced by other pressing matters to attend to during rock climbing? Block (2007; 2011) argues that we can have phenomenal experiences that we cannot access with words and verbal thought. It is first when they are labeled that we can think about them. But does having a conscious experience and thinking about an experience amount to the same? To LeDoux they are the same. But Susan Blackmore arguably needed a source for her recall. Something happened back then. There was something about this episode of her looking at her clock that made it possible to recall. For example, she probably could not have recalled how many steps she had walked, even though she did perform that behavior as well. Explicit memory clearly plays an important role.

Explicit memory, also known as declarative memory, allows for knowledge of semantic and episodic facts that can be consciously and deliberately recalled. This is also referred to as "know-that" knowledge. Semantic facts are such as "apples are red and grow on trees" and episodic facts refer to memories of specific personal events such as "yesterday I ate a tasty red apple." Language, with symbols and concepts, are an integral part of human explicit memory. This allows the subject to put together information from non-current time and place. Explicit learning can be verified by a verbal report. The human storyteller relies on explicit memory.

Memory is closely related to learning. But explicit memory cannot account for all kinds of learning. We also need implicit memory. Implicit memory is involved in learning skills, typically motor- or perceptual tasks. This is referred to as "know-how" knowledge. But it is also involved in cognitive skills, which can become habits in the sense that they become automatized. Implicit memory is also referred to as non-declarative memory because it is information about how to perform something that you cannot verbalize. Implicit memory builds up through repetition, for example you have to practice to ride a bike. Implicit memory has a closer connection to the stimulus and the context present under learning than explicit memory, but is less flexible than explicit learning because it is not independent of time and place. Implicit learning is verified through observation of behavior.

Learning is considered implicit when the subject acquires new information without intending to do so and this resulting knowledge is difficult to express (Berry & Dienes, 1993). LeDoux's view implies that for example the process of learning to ride a bike is nonconscious. A consequence of this is that there is no "implicit self", no experiential center, playing a role during implicit learning. Cleeremans & Jiménez (2002) argue however that implicit learning relies on at least some primitive ability for the cognitive agent to be able to experience the consequences of own behavior. According to their theory, consciousness is a graded phenomenon, differing in degree but not in kind. These primitive experiences are integrated into complex structures that allow for representation of a self, and this constitutes a fully conscious system. On this view these primitive experiences are present without the representation of the self.

Humans can implicitly learn new skills through sensitization, conditioning and instrumental conditioning without having any explicit memories of doing so (Pessiglione et al., 2008). Patients with amnesia can acquire new cognitive skills (Musen& Squire, 1991) and demonstrate complex emotion based learning (Evans-Roberts & Turnbull, 2011). Similar results have been demonstrated in animals. Monkeys as well as rats with damages that result in failure at tasks such as object recognition can retain skill-learning abilities and learn habits (Zola-Morgan & Squire, 1984; Packard et al., 1989). Alzheimer patients can still show emotional learning based on the experience of emotional reward or punishment (Damasio, 1999). They can for example learn to avoid handshakes with people, marked by red labcoats, with tacks hidden inside their hand. The implication of LeDoux's theory is that for example in this last case there is no conscious experience of pain involved in the learning process. Even though the patients demonstrate pain behavior.

Humans learn from pain. LeDoux would argue that painful experience relies on labels that identify nonconscious pain processes. Humans that suffer from genetic disorders that render them insensitive to pain are not able to maneuver in the world without causing considerate damage to their bodies: they bite their tongue, burn their skin and twist, strain and break their limbs (Schon et al., 2018). Even though these patients can learn to have explicit knowledge about what is dangerous: that knives are sharp and that pots are hot, they are not able to avoid damage. This must at least indicate that there is something more about a conscious state than explicit information. The fact that they can have this explicit part without the phenomenal part indicates that these can come apart.

In fact, LeDoux himself refers to examples of explicit learning without implicit learning (Bechara et al., 1995). Given that people lacking explicit knowledge still can learn to avoid painful stimuli, and that people who lack the capacity to feel pain but have the explicit knowledge intact do not, this indicates that the implicit part is doing the explanatory work through phenomenal experience, even possibly rendering the explicit part superfluous in this learning. Based on this it is difficult to explain how animals learn to avoid painful stimuli if they have neither the phenomenal nor the explicit capacities.

Whether or not conscious processes are involved in implicit learning, it is a fact that humans cannot by introspection reveal whether or not they were conscious at the time of the event because our recall and verbal report necessarily involves explicit memory. Susan Blackmore could not determine with certainty whether or not she was conscious when she looked at her clock. This poses a challenge for the human verbal report.

A last point concerns the fact that LeDoux's storyteller is conscious rather than nonconscious. He believes that conscious experience allows us for example, to understand why we are aroused in a particular way; we feel fear when we become consciously aware that the brain has nonconsciously detected danger. He writes: "For once it exists, it opens up the resources of the conscious brain to the quest to survive and thrive" (LeDoux, 2015, p. 20). LeDoux does not however, provide an account of what advantage consciousness may entail; why the same function could not be realized by nonconscious processes alone.

I have argued against the idea that conscious experience in humans requires explicit labels. An explicit concept of self is not necessary for an experiential center. This allows for consciousness in non-linguistic animals hence it is not necessary to pursue the matter of a non-linguistic labeling system.

4.1.2 (2) Neural and cognitive requirements

LeDoux's list of necessary ingredients for consciousness requires specific neural circuits in the neocortex and specific cognitive functions that can realize explicit conscious experience. In the previous section I argued that explicit experience is not necessary for a state to qualify as conscious. Therefore I will not examine what kinds of conditions are necessary to support explicit consciousness. What I will do is see whether or not there is scientific evidence that humans different significantly from other species in terms of neurophysiology.

"Corticocentric myopia" refers to the biased thinking on the role of the human neocortex (Parvizi, 2009). This is based on an underlying assumption that humans are superior to other species, and a misconception that evolution plays out in a specific direction towards better and more advanced products. The task then becomes to find out what accounts for this superiority. Humans look to areas they believe they excel: the neocortex and cognitive functions.

In the search for what physically distinguishes the human brain from other brains many hypotheses have been put forward: brain size, brain size relative to body mass or the size of cerebral cortex relative to brain mass. None of these proved significant. It has now however been demonstrated that brains are packed differently: some brains contain more neurons per unit than others, irrespective of brain or body size (Herculano-Houzel, 2011). The human brain contains the highest absolute number of neurons or connections, and this may explain our cognitive abilities. However, another important finding was that the neuronal scaling that applies to the human brain is the same as the primate brain and great ape brain. This means that the human brain is just a linearly scaled up primate brain when it comes to number of neurons. Based on this there seems to be no qualitative difference, only a quantitative one, in the number of neurons and connections. Homo sapiens has a very short history in evolutionary perspective. If one accepts evolution as the driving force of human design, then it is difficult to see how our species can deviate radically at least in physiological structures, from other species in this short time limit. Consciousness would have to be the result of a sudden incidence.

The case for cognition is similar. The literature on human and animal cognition is permeated by a deep conflict related to the status of humans compared to other animals (de Waal & Ferrari, 2010). One side argues that human cognition is superior, and aim to test whether or not animals have the same abilities such as self-consciousness, theory of mind and language. When animals fail they look for physiological differences that can explain this. This approach has led some to the conclusion that there is functional discontinuity between human and non-human minds (Penn et al., 2008). They seem to assume that recently in evolution something has happened in the human brain that sets it apart from other animals. In their view, investigating human cognition by looking at human ancestors is to be making "naïve evolutionary presuppositions" (Bolhuis & Wynne, 2009). A common assumption is that cognition is closely connected to conscious experience. Therefore, organisms with other cognitive abilities than humans are considered less conscious.

On the other side is the view that humans have evolved like other animals, based on the shaping of basic building blocks, both physiological and functional, and that there are most likely many things in common that we share with other animals that can help us understand our own cognition (deWaal & Ferrari, 2010). What sets the human mind apart from other animals are species specific adaptations just like other species have their specific adaptations. Insects for example, display complex cognitive abilities (Barron & Klein, 2015). There is much support in the literature that indicates empirical continuity between human and nonhuman primate cognition. For example, experiments with macaques on category learning indicate that they have some of the structural components necessary for explicit cognition (Smith et al., 2010). Chimps seem to be able to understand many aspects of other minds such as goals, intentions, perception and knowledge (see review by Call & Tomasello, 2008).

According to Panksepp & Biven (2012) conscious experience arose in subcortical structures. Parvizi (2009) argues that the human obsession with our higher cognitive functions and the role of our cerebral cortex, which is still prevalent in the cognitive neurosciences today, has down-played the role of subcortical mechanisms. Many ideas about the brain that are still alive actually stem from a time where very little was known about the brain. The result is a misconception that there are "higher" structures that dominate "lower" structures, when in fact it may be the other way around. The relationship is at least reciprocal. The cortex cannot be divided from the basal ganglia. The result is that sub-cortical structures are rarely investigated in studies of cognition and most of the technological methods that have been developed are designed to investigate cortical structures, not the subcortical. Merker (2007) argues that children lacking neocortex display facial expressions and behavior that indicate conscious experience.

LeDoux's definition of consciousness is arguably intrinsically tied to what human cognitive abilities and natural language allows for, sometimes referred to as "full-blown conscious experience". His list of requirements for conscious experience is exactly the neural and cognitive factors necessary for these abilities. LeDoux seems that have a preconceived position. Given the fact that the parts and mechanisms of consciousness are still unknown it seems premature, narrow-minded and speciecist to draw such a conclusion.

I have argued that contrary to LeDoux's claim, science seems to indicate that the human brain does not differ essentially in terms of neurophysiology from species that are closely related. The differences may account for the differing cognitive abilities, but there is not reason that these should be essential for consciousness.

4.1.3 (4) Scientific evidence of human consciousness

LeDoux starts out from a default position that processes are nonconscious. He requires rigorous and compelling scientific evidence for consciousness. He knows by direct evidence that he is conscious. He assumes that his conscious experience is closely related to his neurophysiology, both in terms of present parts and in terms of how these parts are related to each other both physically and functionally. Therefore whether or not another organism is conscious depends on how similar that other organism is to him. Other humans are similar in the right way, he believes, and they can verify through verbal report (language) that they are conscious. This verbal report is then evidence of consciousness, but not only that – the fact that they are conscious is evidence that he has found the "right" neurophysiological criteria for consciousness. The verbal report becomes very essential in his theory because this is what ultimately verifies his neurophysiological requirement.

Can the verbal report from another human mind demonstrate that she is conscious according to LeDoux's definition? He requires that the subject demonstrate that they are both explicitly aware of the experience and their self in that experience. I have argued previously that this definition is very narrow. One consequence is that it does not encompass all kinds of conscious experiences, but another quite serious consequence for his theory is that it is difficult for humans to determine by introspection whether or not they were conscious at the time of the actual event. This reduces the reliability of the human verbal report as a correlate and this undermines his theory. Because he cannot know if his physiological criteria constitute the necessary conditions for conscious experience.

4.1.4 (5) Scientific evidence of animal consciousness

LeDoux believes he has evidence for human consciousness. But he lacks this for other species. And this lack is not contingent; not only does science not show us that animals are conscious but in fact it cannot show us. Science cannot show us because animals lack the features that we use as evidence that humans are conscious, i.e. language and the verbal report. Language is both constitutive of and evidence of conscious experience.

I have argued above that his narrow definition reduces the reliability of the verbal report of other human minds. In fact, it can be argued that human verbal reports cannot demonstrate conscious experience according to his definition any more reliably that an animal can through a behavioral substitute for verbal report. For example, there are studies that compare learning in humans and monkeys (Smith et al., 2010). The humans use keys on a keyboard and the monkeys use a joysticks in a categorization task. Can these reports verify that one is explicitly aware and the other not? I argue that no. If the human subject was asked: Were you conscious? She cannot know whether or not the conscious experience originated during recall. LeDoux hence does not have reason to assume that the person can verbally report something that the animal cannot communicate through substitute behavior.

Even if it the case that humans cannot communicate LeDoux' kind of conscious experience, he might object that humans can have these explicit experiences and that these are constitutive for consciousness. I have however previously argued that explicit experience involving labels and hence language, are not constitutive for conscious experience. He has to counter this, but it will still not be possible to demonstrate.

But let us assume then that there were no other human minds, just LeDoux and other species. He did research on himself and his own NCC's and made conclusions about physiological requirements. Then he looked at the other species. Were they similar in the right way for consciousness? What would science tell him? Science would tell him that there are quantitative differences between the human brain and the brains of other close relatives, such as primates, which can account for cognitive differences. However, science would also tell him that there are no qualitative differences that can account for how the human brain can be conscious but not the primate brain (Azevedo et al., 2009). According to science the difference is not essential.

What then about animals further apart from us? What is the cut-off point where a brain becomes qualitatively different? Different species have evolved to solve different problems and therefore brain functions can be realized by other neural structures than in the human brain. This may for example be the case in the insect brain (Barron&Klein, 2015). LeDoux does open for the possibility that consciousness can be realized by different structures: "Even if consciousness is present in other animals in some form, it can't exist in the way made possible by the human brain" (LeDoux, 2015, p.51). On this view then, his theory is only a theory about human consciousness, not a general theory of consciousness. And regarding these species that differ qualitatively in terms of neurophysiology, he can only deny them that consciousness in their organism is realized in the same way as in humans. Human-kind of experiences are likely tied to how they are realized. But we cannot use our kind of experiences and the neurophysiological conditions that allow for these to deny the phenomenon of conscious experience in other species.

LeDoux dismisses behavior as a correlate of consciousness in both humans and animals. In his model, the nonconscious processes that underlie animal and human behavior, are processes that also affect conscious experience. This makes it impossible to know if it is conscious or nonconscious processes doing the causal explanatory work. Science shows us that humans can respond both in physiology and behavior to stimuli they deny seeing. This is why he needs to ask the subject whether or not they were explicitly conscious.

LeDoux model seems to involve an underlying assumption of human superiority. He makes that assumption that because humans can respond nonconsciously this is also the case for all other species. And then he even generalizes this to apply to all behavior in all other species. This underlying assumption is illustrated by his talk where he comments about the laboratory rats and the electrical shock (LeDoux, 2011): he is assuming that humans feel more than animals. If the shock doesn't hurt him then it doesn't hurt the rat. He does not consider that rats could be more sensitive to electrical shock than humans (and the fact that he weighs 280 times as much as the rat).

4.2 Consistency of claims

In claim (4) LeDoux states that scientific observation can inform us about consciousness in other humans. In claim (5) he states that this is not the case for animals. This is based on verbal report and neurophysiology. I have argued that with his narrow definition of consciousness it is challenging to demonstrate human consciousness with a verbal report. In fact this verbal report might not be able to demonstrate more than a behavioral substitute provided by an animal. It is not clear that one can be used as evidence and not the other. Furthermore, I have argued that evidence from science indicates that the human brain does not differ significantly, at least from non-human primate brains, in ways that could account for differences in consciousness. Therefore, (4) and (5) are not consistent.

In his theory of consciousness (2) he claims that the presence of certain neural and cognitive factors is sufficient for consciousness. In (5) he claims that animal consciousness cannot be determined by science. However, the neural factors he requires and possibly also some of the cognitive functions can, at least in principle, be observed by science. Hence these two claims are not consistent. He even goes on to conclude that animals are not conscious based on his scientific theory, and at the same time claiming science cannot tell us anything about consciousness in animals.

Finally, a quick comment to his claims about his lab rats and what they experience: "..the shock is relatively mild…it is not hugely painful to them…we put our hands on the grid all the time to verify that it is working properly" (LeDoux, 2011). This taken together with his other claims: Science cannot tell us about consciousness in animals, we can never know what they feel and it is wrong to assume that their experience is anything like ours, is inconsistent.

4.3 Conclusion

I have argued that labels, and hence language, are not constitutive for human conscious experience. An explicit concept of self is not necessary for an experiential center. This allows non-linguistic animals to have conscious experiences. There are two challenges with LeDoux's narrow definition of consciousness. First, it does not capture all kinds of human conscious experiences hence it may exclude conscious experience in humans as well as animals. Secondly, it makes it challenging to demonstrate consciousness in any subject, also humans. In fact I argue that a human verbal report cannot demonstrate this kind of consciousness any better than a substitute report by an animal. Therefore, if a verbal report can be used as a correlate in humans, then the same is the case for animals.

I have also argued against LeDoux's claim that humans and animals differ significantly in terms of neurophysiology. His theory requires specific neural and cognitive resources that allow for explicit knowledge. Although human brains differ quantitatively from non-human primate brains in ways that may account for differing cognitive abilities, there is no scientific evidence that human brains differ qualitatively in a way that would make consciousness exclusively human. Therefore, if neurophysiological correlates can be used for humans, they can also be used for animals. I have also argued that LeDoux's neurophysiological requirements pertain only to how conscious experience can be realized in humans and other animals that do not differ qualitatively from them. He can however not deny that consciousness may be realized differently in other species that differ qualitatively from humans.

Several of LeDoux's claims are inconsistent. First, it is inconsistent to claim that scientific correlates can tell us about consciousness in humans but not animals. Second, he argues both that science cannot tell us about consciousness in animals when in fact he has a neuroscientific model with neurophysiological requirements. The presence of these can in principle be demonstrated; hence according to his own theory neuroscience *can* tell us about consciousness in animals.

I have concluded, based on LeDoux's definition of consciousness: *if* correlates of consciousness, the verbal report and neurophysiology, can tell us about consciousness in humans *then* they can tell us about consciousness in animals. In the next section I will look at behavior and what this can tell us about consciousness.

5 Presentation Claim 2

LeDoux believes that he can provide an exhaustive explanation of animal behavior in terms of neuroscience, without referring to consciousness. We should not postulate feelings such as fear into the system when it is not necessary for the explanation. He refers to Morgan's Canon when he states the following: "I believe we should not assume conscious feelings in animals if nonconscious processes can account for the behavioral effects" (LeDoux, 2015, p. 129). In this section I will present how LeDoux describes and explains animal behavior. His case in point is how animals detect and respond to danger. Learning is closely related to behavior and I will present his view on how learning can be explained by nonconscious processes. I will discuss this material in the following section.

5.1 Detection and response to threat and danger

5.1.1 Background

At the time LeDoux started his research, the common view regarding the relation between behavior, consciousness and evolution was in short the following: Reptilian ancestors were believed to behave based on reflexes and instincts, elicited by automatic and nonconscious processes. Later the mammals evolved with a new brain system, the limbic system, which caused emotions (consciousness) to emerge in the organisms. This made behavior more adaptive. Further out in the evolution of mammals, the neocortex emerged and this allowed for emotional control and thinking. According to this limbic-system theory, emotions - the capacity for consciousness - had evolved along with subcortical brain areas. This view still has many supporters today (Panksepp, 2011; The Cambridge Declaration on Consciousness, 2012).

The amygdala, two almond-shaped groups of neurons that are part of the limbic system, has been an important area of research for LeDoux for decades[2]. The amygdala plays an important role in the processing of threats and danger, and has been assumed to play an important role for conscious emotion. When LeDoux entered the arena he observed that as neuroscientific research increased, more and more brain areas were assumed to be involved in emotion. He started questioning the theory, and the role of the amygdala for emotion and consciousness: Could it be that activity in the amygdala could nonconsciously both process emotional stimuli and cause behavior (LeDoux, 1998)? LeDoux felt this warranted for a new theory; it was time to start fresh, free from the earlier assumptions.

LeDoux's thinking is influenced by his early graduate work with Michael Gazzaniga on split-brain patients (Gazzaniga & LeDoux, 1978). They had studied how information travels in these patients' brains: from stimulus to output behavior. In split-brain patients, the connection between the right- and left-brain hemisphere, the corpus callosum, is severed to reduce epileptic activity. The result is that one side does not have access to information about the other side. This means that if an object is presented to only one hemisphere (for example one eye), the other side does not have information about it. Usually the left hemisphere is responsible for processing of language and therefore communication with the right hemisphere is difficult. At the time it had even been argued that the right hemisphere was not conscious (Eccles, 1965).

There was however, one young male patient in their study that had developed some language abilities in the right side as well, due to childhood physical trauma, and this side was therefore able to process language to a certain degree. This allowed for communication with the right hemisphere, which yielded some intriguing results. For example, the two hemispheres sometimes gave different answers to the same question, possibly indicating two minds. But what surprised LeDoux even more was that if the right hemisphere was presented with an emotionally loaded command such as "Kiss"

[2] He even has a band called "The Amygdaloids"

(Gazzaniga & LeDoux, 1978, p. 151) this resulted in a verbal outburst from the left hemisphere of "No way!'" The patient however, did not know why he had said this because he did not have access to the command. When the stimulus was presented to the other side the reaction was the same, but now he could refer to the command: "No way, I'm not going to kiss you guys." The patient was a callosum-sectioned patient with his anterior commissure spared. Thus the inter-hemispheric limbic connections were intact. These limbic structures were assumed to play a role in emotion, and the information was likely transferred through these. But the patient's lack of knowledge about the object (the command), and still having an emotional response was interpreted by LeDoux to indicate that emotional and perceptual information is processed separately in the brain.

Another important observation was that the patients would make up stories that fit with their behavior. For example, a patient could be induced to wave their hand through the right hemisphere. When they asked the left hemisphere of a subject why he had waved his hand he would for example answer: "Oh, I thought I saw someone I know." Similarly, a female subject was shown nude pictures and started giggling. When asked why, she replied: "I laughed because the machine is funny." This gave him the idea that the function of conscious experience is to interpret physiological states and behavior and in this way work as a storyteller.

These findings on split-brain patients gave rise to LeDoux's current ideas on how consciousness works and its function: Organisms can detect and process stimuli nonconsciously, and that conscious experience arises when an organism is able to self-represent and interpret these events. At that time however, he had not yet developed these thoughts. He still assumed that non-human animals, like humans, had conscious emotions. His main interest was to understand how an organism learns to relate emotional value to an object or event. His plan was to follow how information traveled from stimulus to sensory system to muscular system and behavior in order to understand how emotional significance was added to an object as the information traveled through the brain. These brain mechanisms were only possible to study in animals so he started experimenting on rats.

LeDoux's main focus is to understand the processes that underlie the emotions fear and anxiety in humans. His animal research therefore centers on behavioral and physiological responses to threat and danger.

5.1.2 Nonconscious physiological processes

The animal brain consists of many different neural circuits that operate to ensure survival and reproduction. These control such things as internal physiological processes, reproduction, foraging behavior as well as detection and response to threats and danger. LeDoux calls these latter for defensive survival circuits. Animals have innate, automatic systems that prevent them from being harmed or eaten by predators. Predation is likely to be an important source for the development of circuits underlying fear and anxiety. Typical defense reactions are freeze, flight or fight. Freezing helps to avoid detection, flight gets you away – and if it is too late for either you fight.

Physiological responses and behavioral responses are closely linked. LeDoux reminds us of the myriad of often overlooked, nonconscious physiological processes that are involved in generating behavior. These have to engage and prepare quickly: turn on and off the right processes and distribute energy to the right tissues and organs. The autonomic nervous system, consisting of both the sympathetic- and parasympathetic system, plays an important role: adrenaline acting within seconds, and cortisol in minutes or hours. These nonconscious physiological processes are part of a causal chain in the production of emergency behaviors, both innate and learned. LeDoux believes that animals have evolved to detect and respond to threat and danger nonconsciously. They need not feel either fear or anxiety in order to behave adequately. Humans have inherited these nonconscious processes from our evolutionary ancestors. But, humans have not inherited emotions from our ancestors. Humans have specific circuits and cognitive abilities that can interpret these nonconscious processes consciously.

5.1.3 Stimulus, "central state" and response

LeDoux has a model (LeDoux, 2015, figure 2.7, p. 45) that illustrates the relation between stimulus, consciousness and response in a threat situation. A stimulus will elicit what he refers to as a "central state." Such as state has been central in many models of behavior, where some involve consciousness and some don't. The behaviorists treated this as a nonconscious state. As consciousness could not be measured, they instead measured behavior and explained it in terms of nonconscious processes. Fear, on their view, was not a feeling but rather this relation between stimuli and response. Later, in behavioral psychology, fear was considered a central motivational state, but also here nonconscious. The idea was that activation of the defensive survival circuits caused the central motivational state, which in turn caused the behavioral responses.

There are also theories that regard this state as conscious. In what LeDoux refers to as "the commonsense view", this state refers to fear as an emotional state of mind, which plays a causal role for eliciting fear behavior. Darwin (1872) was a proponent for this and believed we had inherited this from our ancestors. This is the common intuitive human interpretation of the role of the feeling fear: the stimulus causes fear that in turn causes behavior. This is in accordance with basic emotions theory (Panksepp, 2011). On this view, emotions are hardwired in the brain. When the animal is exposed to a threat this triggers survival circuits and this in turn puts the whole brain in a certain state, which is the neural instantiation of the feeling of fear, which then causes defense responses.

LeDoux's view is a version of central state, but with some important differences. First of all, to avoid confusion he no longer operates with the term "fear" in his models. Instead of calling the amygdala circuits for "fear-circuits", he now calls them "circuits that detect and respond to threat." This way he demonstrates that these are nonconscious processes. What distinguishes his model from the others is that activation of defensive survival circuits triggers both a defensive motivational state throughout the brain *and* the defensive responses. The state is thus a result itself rather than a cause for the responses. This defensive motivational state contributes to motivation of instrumental behaviors. Defensive motivational states are present in both complex and simpler animals, but they cannot experience these states. In humans the central motivational state provides neural ingredients into higher cognitive systems, such as attention and working memory, which can interpret this and in this way contribute to the feeling of fear. Feelings are psychological constructions.

5.1.4 Examples of nonconscious detection and response

LeDoux' refers to human examples where subjects deny seeing stimuli, but still respond behaviorally and physiologically: "...threat processing does not require consciousness. It is fundamentally an a-noetic (implicit or nonconscious) form of processing" (LeDoux, 2015, p. 205). As previously mentioned, this is what makes it challenging to use these as correlates. I will present his interpretation of these examples more closely.

In blindsight, the subject has damage to the brain that renders certain parts of their visual field blind. The subject will therefore have no conscious awareness of what is in this blind area and hence will deny seeing anything there. They can however still be coaxed to respond to stimuli (Milner & Goodale, 2006; Stoerig & Cowey, 2007). Threatening stimuli presented to the blind area can also cause responses in their autonomic nervous system. LeDoux takes this to mean that the brain processes these stimuli even though the subject cannot consciously see them. He explains their ability to act on these "unseen" stimuli: the brain processes a lot of information about the stimulus, not only the identity of the stimulus, but also such things as location and movement. If the system responsible for identifying the object is damaged, then the subject will not be able to identify the object and will hence deny having seen it - even though they can respond towards it based on other kinds of information processed nonconsciously. Identification is necessary for the subject to be explicitly aware and hence conscious of the object, but the rest is all nonconscious processes.

In visual neglect subjects suffer from damage in areas related to attention, and not the visual cortex. The result is that the subject is not able to direct attention to the object and hence cannot report consciously seeing the stimulus. In one such patient, brain imaging has shown that the part of visual cortex involved in processing faces is active when the patient is shown pictures of faces, even though the patient denies seeing them. Block (2007), who distinguishes between phenomenal and access consciousness, interprets this to mean that the patient was phenomenally conscious of the faces, but was not able to access this experience and verbally report it due to lack of attention on the face. LeDoux denies this phenomenal experience and instead suggest that lack of attention prevents the subject from explicitly experiencing the stimulus.

There are also examples with normal brains that demonstrate response to stimuli not consciously perceived. In many situations, the reaction is simply too fast for conscious experience to have time to unfold and play a causal role for behavior. According to LeDoux the circuitry involved in such cases are parts of a "low road", which is a faster route that handles less information, and allows for fast responses. For example, when a person sees a snake in the grass and jumps away. In this case the experience of fear does not play a role for the execution of behavior. This shows that visual systems can operate nonconsciously, process the stimuli in a meaningful way and elicit adequate behavior. But there are also complex nonconscious cognitive processes that can affect behavior. For example, in the case of human decision-making based on intuition, heuristics are used: fast nonconscious strategies, but the subject cannot explain the reason for their decision (Kahneman, 2013).

Similar results can be obtained in subjects with normal brains manipulated in experimental settings. By presenting stimuli subliminally, i.e. presenting stimuli too fast for conscious processing or by masking a stimulus by presenting another stimulus immediately afterwards, the subject is prevented from consciously seeing the stimulus. Such studies allow the researchers to compare which areas in the brain are active during the time when the subjects claim to see and when they deny seeing anything. The results show that consciously seeing stimuli correlates with activity in certain areas of the visual cortex and areas associated with working memory and attention (Dehaene, S. & Naccache, L., 2001), but when the subjects deny seeing things, only the visual cortex is active. Similar results have been found for other sensory modalities such as auditory. The relation between what the subject sees and which brain areas area active are the same as the findings in subjects with blindsight.

What all these examples have in common is that there is something (blindsight, visual neglect, masking) that prevents access to the systems that are involved in identifying the stimulus and allow for conscious processing. And this is the system that makes humans conscious. But as LeDoux argues, there are plenty of other systems at work that can account for behavior. These examples then, together with the examples of split-brain patients who interpret their own behavior, is taken to argue that a lot of behavior can be explained without consciousness. Humans and animals have a long common evolutionary history. If this is the case for humans, then he believes it is the case for animals. From this he infers that this is the case for all animal behavior.

5.2 Learning

An important feature of animal behavior is learning. This allows them to adequately adapt their behavior. An explanation of behavior needs to be able to account for this. This section describes how LeDoux explains learning. I have also included some information about learning from Kandel et al. (2000) that he does not explicitly state in this book, but which he builds upon. I believe this is relevant for the discussion part that follows later.

5.2.1 Learning and behavior

Animals live in a dangerous world with both predictable and unpredictable stimuli. Some dangerous stimuli are always present in some form or other. Behavioral responses to these are likely to be innate. Other dangers vary between generations or through the lifespan of an individual and it is advantageous to be able to respond to these successfully.

Behavior has traditionally been divided into reflexes, instincts and various kinds of learned behavior. Reflexes are considered elementary responses: innate species-specific, automatic and involuntary, and in that sense nonconscious. When a human freezes at the sight of a snake or a rat freezes to electrical shock, these are reflexive behaviors. Instincts are more complex patterns of behavior. Both reflexes and instincts have been considered to be innate in the sense that they are behaviors that the animal is wired to perform and does not have to learn. LeDoux argues however that behavior is plastic in the sense that it is difficult to clearly distinguish between what is innate and what is learned behavior (LeDoux, 2015, p. 28). The behavioral plasticity varies between different species. Some species rely on simple reflexes to keep them alive while others develop elaborate behavioral patterns. But even the animals that exhibit the ability to learn complex tasks also rely on simple reflexes. Animals with more complex nervous systems tend to perform more learned behavior. It is now also widely accepted that the expression of genetic programming is affected by individual experience.

5.2.2 Learning at the behavioral level

Every time an animal performs any kind of behavior it arguably involves learning: elements of behavior are either increased or decreased in frequency. Even simple organisms can learn and successfully modify their reflex behavior (Kandel et al., 2000). There are two ways to increase the frequency of a behavior: The presence of an appetitive stimulus acts as a positive reinforcer and removal of an aversive stimulus acts as a negative reinforcer. Similarly there are two ways to reduce the frequency of a behavior: removal of an appetitive stimulus acts as a positive punisher or adding an aversive stimulus, which then acts as a negative punisher.

Learning is commonly divided into two main categories: non-associative and associative learning. In non- associative learning, repeated exposure to a stimulus can lead to changes in response frequency of response. Through the process of habituation it diminishes and through the process of sensitization it increases. A reflex behavior can be modified through these processes.

A reflex behavior can also be subject to classical conditioning, also referred to as Pavlovian conditioning. This is a more complex form of learning than sensitization, in that the organism learns to associate a new stimulus with a reinforcing unconditioned stimulus. The animal does not learn new behavior, rather it learns to elicit an established behavior in the presence of the new cue. Learning to form associations between stimuli allows the organism to identify and remember cues in the environment that precede the stimulus and in this way it can avoid the encounter with the dangerous stimulus. LeDoux uses threat-conditioning in his lab with rats: the animals learn to associate a conditioned stimulus (CS) such as a tone with an unconditioned stimulus (US) such as electrical foot shock from the floor. The tone then becomes a predictor and elicitor of the reflex behavior. The rats typically freeze as a response to the tone. Not only the tone, but the context as well plays a role. An animal has to figure out which part of the environment, or the context, is a predictor of danger. This can be anything the animal can sense: a smell, sound, an animal, an object, an event, a place, a situation etc. For example, the chamber or cage itself can figure as a CS.

During instrumental conditioning, or reinforcement learning, the animal learns to associate specific motor patterns with reward or punishment. In addition to modulating a response, this allows for the creation of new behaviors. Motor patterns and combinations that result in reward have an increased likelihood of being repeated, whereas behaviors that result in punishment are less likely to be repeated. The animal does not have to know how much each action in the sequence contributes to the outcome. This is also referred to as "The law of effect" (Thorndike, 1898). A typical experiment in LeDoux's lab, which produces avoidance-behavior is the following: A rat is put into a box with two compartments. In one of these the rat is threat conditioned with a tone and shock to the feet delivered through the floor. This induces the rat to freeze at the sound of the tone. If this is repeated over and over again the rat begins to make movements based on innate responses, previous learning or random activity, and eventually it enters the other compartment and in this area there is no shock. After several trials, the rat learns that it can escape the shock by running to the other compartment. Now the CS (the tone), by virtue of the relation between the CS and the US, becomes a stimulus that motivates the rat to perform this avoidance behavior.

Non-associative learning allows the animal to adjust its response to the stimulus. Classical conditioning allows the animal to predict and therefore avoid the unconditioned stimulus. Instrumental conditioning, allows the animal to both adjust responses and create new ones. The animal uses previous experience to improve behavior.

5.2.3 The neuroscience of learning

How does learning described at the behavioral level translate into the neuroscientific level? Animals with a complex nervous system, such as the rat, have complex systems of interneurons and are thus difficult to study and describe. Eric Kandel, a pioneer in the field of learning and memory, turned to an organism so simple that he was able to describe its whole neural system: the sea slug *Aplysia californica*. This organism has a nervous system consisting of only about 20 000 central nerve cells and has defensive reflexes that consist in withdrawal of its gill and its syphon (see Kandel et al., 2000). These can be habituated, sensitized and classically conditioned. He was able to describe the events that occur in the learning processes at the cellular level.

The common factor in all learning processes is changes in neuron connectivity, i.e. the modification of synapses. This changes the efficiency of the synapses that are involved in the pathways that execute behavior. Kandel was able to determine the cells, the synapses and the molecular mechanisms involved in the cellular change that takes place during learning the sea slug. I will make a short presentation of Kandel's description because they are necessary to understand LeDoux's description of reinforcement.

In habituation, stimulation of the sensory neurons with a novel stimulus generates excitatory synaptic potentials in the interneurons and motor cells. If the stimulus is non-harmful, repeated exposure results in decreased transmitter release in the presynaptic terminals of the sensory neurons. This reduces synaptic transmission between the sensory and motor neurons, as well as between certain interneurons and motor cells. The result is that the motor neuron fires less briskly. These changes in synaptic strength occur at several places in the reflex circuit and the memory is thus distributed through the circuit, not in one specific site.

The cellular events that occur during sensitization are similar, but the result is presynaptic facilitation and the response is increased. This is a generalized response in the sense that stimuli that cause sensitization arouse the animal and make it more likely to respond more vigorously to consecutive stimuli, even though they are not necessarily dangerous. This is similar to the startle response in humans (Kandel, pers. comm., 2017) The same synapses can be involved in both habituation and sensitization, and hence be part of different memories.

Classical conditioning is an advanced form of sensitization (Kandel et al., 2000; Antonov et al., 2003), an elaboration of a presynaptic cellular mechanism. When the US follows immediately after the CS, the postsynaptic cell is still active when the presynaptic inputs arrive. In other words the weak stimulus' (CS) ability to activate a neuron is enhanced because it co-occurs with a strong stimulus (US) that activates the same neuron. This is caused by molecular changes in the pre- and post neuron. This is also referred to as Hebbian learning, and has been confirmed in LeDoux' lab and is supported by recent studies using optogenetics. Classical conditioning and instrumental conditioning operate by similar laws and thus might use the same underlying neural mechanisms (Kandel et al., 2000, p. 1242). It is therefore likely that instrumental conditioning can be accounted for in the same way as classical conditioning.

5.2.4 Reinforcement

Learning at the behavioral level involves reinforcement of behavior. At the neural level there is synaptic modification. Can reinforcement by explained in terms of these neuroscientific events? LeDoux notes that over history there has been both a conceptual change and a shift in the terminology regarding reinforcement. First the idea was that hedonistic subjective feelings guide the animal, i.e. animals behaved in order to avoid pain and obtain pleasure. This idea was challenged in the 1920s by behaviorists who argued that science should focus on objective measures of behavior. The subjective terms were then replaced by the term “reinforcement” and the focus went from the subjective experience to become properties of the stimulus.

Reinforcement is the tendency for environmental stimuli to strengthen learned stimulus-response tendencies (White, 1989). How this can be explained will depend on whether or not the subjective feelings or the stimulus is assumed to be doing the explanatory work. Hedonistic theories explain it terms of the subjective feelings: positive feelings from reward and negative feelings from punishment such as pain. Panksepp (Panksepp, 2011; Panksepp & Biven, 2012) believes there are basic primary affective emotions or states such as fear, rage, panic and lust and that these are present in all mammals. These emotions, such as fear, are realized by the same circuits that elicit behavior and therefore play an important role in reinforcing behavior, rather than the execution or control of the behavior. The animal learns the meaning of the tone, the tone induces fear and the animal instrumentally behaves to reduce this. What is reinforced is the relief from the fear-state (Mowrer-Miller theory, Mowrer & Lamoreaux, 1946).

LeDoux does not agree with this view: “I question the value of viewing brain states elicited by threats as subjective feelings” (LeDoux, 2015, p. 74). These are not feelings themselves, even though they do contribute to the construction of feelings in humans. LeDoux believes that introducing subjective feelings into the explanation of animal behavior raises more questions than it solves. Instead of looking at feelings LeDoux looks to the cellular and molecular levels to understand the source of reinforcement: (LeDoux, 2015, p. 74-75):

> The reinforcement, in my view, is not due to fear reduction but to a reduction of components of the nonconscious defensive motivational state that the CS triggers. That is, behaviors that eliminate the CS are reinforced because the CS is no longer activating the defensive survival circuit, and this, amongst other things, changes the level of neuromodulators, which are known to be important as reinforcement signals in survival circuits and instrumental action control areas.

The important neural changes occur at the synapses in specific pathways and the synapses are regulated by neuromodulators. This is similar to what takes places in appetitive conditioning with reinforcers such as food, sex and drugs. According to LeDoux's studies, the amygdala is involved in both positive and negative reinforcement and may play a role for processing of value in the brain (Morrison & Salzman, 2010). The amygdala is involved in triggering arousal by releasing neuromodulators that mobilize and energize the brain and the body, resulting in increased attention, vigilance and sensitivity to other inputs. The amygdala plays a fundamental role in threat conditioning and extinction in people (Bechara et al., 1995), independently of whether or not the subject is conscious of the stimulus (Morris et al., 1998). Once an avoidance response has become a habit, the amygdala is no longer necessary. But this is all through nonconscious processes.

5.3 Conclusion

LeDoux uses responses to threat and danger as a case in point for behavior. He argues that animal behavior is elicited by preceding nonconscious processes. He presents his model of the relation between stimuli, consciousness and behavior: stimuli elicit both the motivational state of the animal *and* the responses (behavior). So both human and animal behavior can behave nonconsciously, but in humans the motivational state is input to conscious systems that interpret some of these nonconscious processes. This model then explains how some human behavior has been demonstrated to occur nonconsciously.

Learning is essential and he explains reinforcement at the behavioral level in terms of cellular and molecular events: the behavior reduces the input to the motivational state. When an animal behaves in a way that reduces the presence of the dangerous stimulus, there is less input to the motivations state and this changes the level of neuromodulators.

In the next section I will discuss what his neuroscientific explanation can tell us about consciousness. Then I will critically examine his model of the relation between stimuli, consciousness and behavior.

6 Discussion Claim 2

LeDoux claims that he can explain animal behavior in terms of neuroscience. He argues that because he does not need to refer to consciousness, consciousness is not necessary to explain behavior and therefore animals are not conscious. I begin by making some preliminary points about scientific explanations in general and what they can tell us about phenomena such as behavior and consciousness. Then I will look at LeDoux's explanation of behavior and his model of the relationship between stimuli, consciousness and behavior. I go on to identify two underlying assumptions, object to these and see how this affects his examples of nonconscious behavior. I explore learning and reinforcement and how this can be realized in artificial and natural systems. Finally I argue for a new correlate of consciousness.

6.1 Explanations

LeDoux claims that animal behavior can be explained in terms of neuroscience. Human behavior is explained in terms of neuroscience *and* consciousness, i.e. the storyteller. According to his theory however, consciousness itself can be explained in terms of neuroscience. Hence a first immediate response to his claim is that *both* animal and human behavior can be explained by neuroscience. Similarly, a physicalist can explain behavior in terms of the physical. By his line of argument this renders humans nonconscious as well. But this conflicts with his view of human consciousness and is thus an undesirable consequence.

His claim raises some challenging questions. Can an explanation based on one model provide an exhaustive explanation of a phenomenon in such a way that one can conclude that all factors outside this model play no causal role for the observed phenomenon? For example, he believes that he can explain animal behavior in terms of nonconscious cellular and molecular processes, and that this warrants the conclusion that consciousness does not play a role for behavior. If the answer to this is yes, then the question that naturally follow is, is his explanation of behavior exhaustive?

I will now briefly consider explanations in general. Phenomena in the world, whether objects, events or processes, can be categorized, investigated, described and explained in many different ways. They can be characterized from different perspectives, by different models and hierarchical levels. This raises the question: Is there one right way to characterize and explain a phenomenon? A closer look at explanations can be clarifying and help us identify tensions in the debate. The nature of explanations is in itself a large topic and I do not intend to present an exhaustive discussion here. The considerations I will put forward are in line with Dennett's (2017) pragmatic naturalism about explanation in science.

6.1.1 Factors in explanations

6.1.1.1 Different stances

A phenomenon, such as behavior or consciousness, can be investigated from different perspectives. Dennett (2017) describes different "stances" or strategies that can be adopted: the physical stance, the intentional stance and the design stance. The physical stance investigates the phenomenon as a physical phenomenon and explains it in terms of physical parts and causal relations between these. In this explanation behavior can be investigated as a physical event that can be explained in terms of previous physical events. The intentional stance understands the phenomenon in terms of a rational agent with propositional attitudes such as intentions, motivations, desires and beliefs about the world and beliefs about consequences of behavior. An intentional, or mental explanation, will therefore involve the psychological experience of the subject that precede the behavior. The behavior will be considered an action rather than an event. The design stance assumes that the phenomenon has been designed for a purpose and explains the phenomenon in terms of its function, i.e. what the animal obtains by performing the behavior and the adaptive advantage. These stances may have different pragmatic value depending on the explanandum.

The stances involve different explanatory lines, and within these there are different models of the phenomena that in turn involve different parts, mechanisms and underlying assumptions. These commonly involve different kinds of vocabulary. Vocabularies arise to be able to track specific patterns. Humans have the ability to track mental patterns, but this ability to track mental patterns does not mean that anything new has entered at the physical or neuroscientific level. It just means that it can be described by us in a different way using a different vocabulary.

Both the physical stance and the intentional stance can be expressed in terms of preceding events and give proximate reasons for behavior, similar to what Dennett (2017) calls answers to "how come" questions and the proximate questions in ethology. The design stance points to the future and is similar to the "why" question (Dennett, 2017) and the ultimate question in ethology.

6.1.1.2 The physical perspective

The idea that we can have a physical stance rests on the assumption that we can have objective knowledge about the world, even though we perceive the world, categorize it and entertain these categories from our own subjective perspective. We believe that human thinking can abstract away from our subjective perspective and give objective descriptions and explanations, which are centerless (Tartaglia, 2016) with a view from nowhere (Nagel, 1986). We use language and its labeling system as a tool for expressing this knowledge. The common default assumption about the external world is that it is nonconscious[3].

Humans categorize the physical world based on how it appears to us through our perceptual and cognitive abilities. We typically divide nature into disciplines such as biology, chemistry and physics. These different scientific disciplines have different approaches to modeling the world. They involve different parts and mechanisms. The relationship between these models can be difficult to depict, even when they concern the same object or phenomenon. There may be contact surfaces, but the causal relation or line between them can be complicated.

Further complicating this is that within disciplines there are different explanatory levels (Craver and Darden, 2013). An example is neuroscience, which is the study of the nervous system. This is one particular way of looking at the brain. The nervous system can be studied in terms of looking at individual neurons and which parts and mechanisms are causally involved in firing action potentials. One can also look at how neurons affect each other in terms of the synapses that connect them. The nervous system can also be studied at the level of a network of neurons by looking at the transmission of signals and whether they are excitatory or inhibitory. These signals can further be modeled as computational networks. Studying the path of a signal at this level does however not provide information about what caused the individual neurons in the circuit to fire. The signal itself can be realized in multiple ways. There are thus different hierarchical abstractions that may not be directly translated. Some may be reducible to another explanatory level, yet others may not. Similarly, computational science involves abstract hierarchical levels and switching between levels demands thorough considerations (Ballard, 2015). New parts and mechanism are involved, and it can be very complicated to understand how changes at one level affect changes at another level. The nervous system can also be studied in terms of how it is involved in the control of many important functions, such as generation of behavior.

[3] The hard question seems to take for granted that what needs explaining is how to get from physical to mental. It may be that rather than assuming that the world is nonconscious by default, we should assume the opposite and have to explain how there can be nonconscious material in the world.

The physical parts in a model can be observed, as well as their movements and interactions. But we also have models that involve abstract parts such as concepts and even functions. Both “behavior” and “consciousness” are labels for concepts. Even though each of these refer to groups of particular instances, each of these instances will differ, and when these non-physical parts figure in models they are not measurable in the same way as physical parts. It is difficult to depict the relation between such models and the levels. For example, it is challenging to see how physics can explain functions. Even though we today have knowledge about all the neurons and their connections in such organisms as the Aplysia and Nematods, we can still not explain their behavior. This illustrates the complexity of the systems that underlie behavior.

6.1.1.3 The behavioral level

At the behavioral level what is categorized and measured is observable behavior. According to radical behaviorism and Skinner (1974), everything an animal does is considered behavior; this also includes verbal behavior and internal mental phenomena such as feelings and thoughts. But even though inner mental events are also physiology, just like external behavior, inner mental events are not valid explanations of behavior because these events are private and not observable by others. Only external behavior is observable by others and can give explanations. External behaviors are considered events and these events can be studied by the natural sciences. Skinner’s view is also termed a “molecular” view in the sense that he believed that behavior can be explained in terms of atomistic parts of molecules. It is these parts that operate on the environment. Just like the environment in turn operates on the animal. Radical behaviorism is according to Skinner conceptual analysis of behavior, which he viewed as the philosophy of the analysis of behavior.

Behavior can be explained both in terms of proximate events, i.e. the physical events that precede the behavior, and in terms of ultimate events, which refer to the function of the behavior. The relation between the physical perspective, and the functional perspective is difficult to depict. For example, a function relies on neural organization to realize the necessary events. Genetic codes are transferred from one generation to the next. These codes determine the phenotype, which includes anything that is manifested and observable in an organism, such as the way it looks and behaves. Genes codes for neural organization in an organism, which in turn determines behavior which in turn affects behavioral success in terms of survival and reproduction. But it is not the genes themselves that are selected by natural selection, but the behaviors they code for, such as superior camouflage or ability to find food and mates compared to conspecifics. From this perspective, one can argue that the functional design can act on the neural organization (Johnston, 1999).

6.1.1.4 The mental perspective

Behavior can be explained in terms of mental states: the psychological conscious experience of the subject. This is based on the assumption that these mental states are part of a causal line that affects behavior. These mental states are often referred to as ordinary folk psychology or propositional attitudes such as beliefs, hopes or desires. For example, when I go to the train station I do so because I want to travel somewhere and I believe there is a train going there. When I drink water I do so because I am thirsty and I believe that drinking water will relieve my thirst. LeDoux's storyteller is a causal line of such mental states. The storyteller uses labels on experiences and allows the subject to understand its own behavior, analyze previous events, imagine scenarios, and importantly to predict the behavior of others. LeDoux believes that we need to refer to this storyteller to explain human behavior.

Behaviorists however would claim that these mental states are postulated entities and that human behavior can be explained in terms of conditioning instead. According to their interpretation, humans learn these relations and act only based on a myriad of conditioned behaviors. The belief is only an interpretation of the verbal-mind, of reward and punishment. But they would claim that we do not need this belief for the conditioning to occur and hence to explain the behavior. Stimuli that reinforce are mostly conditioned stimuli. An example is pressing the light switch to turn on the light: there are many conditioners such as the sound when you press, tactile stimuli and the visual image of pressing the switch.

6.1.2 One phenomenon – many explanations

Each of the stances, or perspectives on the world, contains knowledge about the phenomenon at hand. They each constitute different models. Each model of a phenomenon will have its own characteristics: its own parts, relations between these parts and arguably even its own interpretation, a specific aspect, of the phenomenon it concerns. But the model must be able to account for the relation between all its parts. One cannot postulate parts. One view is that a model does not warrant conclusion about parts that cannot be accounted for, that are not related to the other parts in the model. The conclusion is that it does not play a role in this model, but it may still play a role for the phenomenon at hand in another model of the phenomenon, hence one cannot conclude that it is not related to the phenomena at hand. This limits the explanatory power of each model. On this view, each model is its own puzzle to be solved. But the different models may also be viewed each as a piece of a larger puzzle.

6.1.3 Consequences for LeDoux

I have discussed various factors that play a role in explanations. Now I want to apply the perspective I have sketched to the issue at hand: What are the implications for LeDoux's claims? LeDoux may not agree with this view of explanations and argue that all the necessary parts and relations are accounted for in his model, and because consciousness isn't a necessary part, this allows him to conclude that it plays no role for behavior. Consciousness thus should not be postulated into the model. He could for example compare an animal to a robot and say that he can explain everything about the robot's behavior based on its parts and programs. Does this not warrant exclusion of consciousness?

This matter might not be as straightforward as he believes it to be. LeDoux's default assumption is that neuroscientific processes are nonconscious. Any allegation of consciousness requires argument. He believes that the arguments for human consciousness, through verbal report and physiological similarity to himself, are sufficient for this and therefore warrants the intentional stance. He claims that animal behavior can be explained by the physical stance, and maybe he is also willing to accept that this is also the case for human behavior. But in the human case he still appeals to the intentional stance, because it is necessary to account for the storyteller. He has a theory of how consciousness arises and animals do not have the necessary parts. So he already knows that they are not conscious. His explanation of behavior, if it is plausible, does not in itself warrant exclusion of consciousness, but rather can be considered evidence of this theory. It can be argued that he is using skepticism selectively. He cannot base this on science because biological knowledge points towards continuity in the animal kingdom.

His theory of consciousness is important for his argument. Because he claims that he can show that consciousness arises from ingredients that are present in humans, but not in animals. I will examine LeDoux's theory of consciousness to see whether or not this provides a plausible account of consciousness.

6.1.3.1 LeDoux's theory

LeDoux wants to explain how consciousness arises from non-conscious parts. In his model he brings together the physical and the mental lines. The parts in his model are the neuroscientific ingredients and conscious mental states. Does he succeed in causally relating these parts? He compares the emergence of consciousness from non-conscious ingredients to how the flavor of a soup can emerge from non-soup ingredients such as water, chicken, salt and carrots (LeDoux, 2015, p. 228): When all the ingredients come together consciousness emerges, and the quality and intensity of the conscious experiences will vary with the ingredients, just like the taste of the soup.

It can be argued against this that his non-soup ingredients already have taste properties, hence it seems the ingredients already possess what he claims to emerge. If the ingredients didn't possess taste, his soup would have no taste no matter how much he cooked it. He has just shown that parts with taste properties can come together to make new taste properties. Just like visually a chair can be made from non-chair parts and as a whole this chair has other visual properties. Or a melody can be made from non-melody parts and have different properties such as minor or major key. But perceptible properties is not something that arises here, it is already present in the parts.

It may be though that what he intended to show is that soup as an object-category can be created from non-soup parts, just like a car can consists of non-car parts, and that this object has new properties such as being able to drive. Likewise, when certain neuroscientific parts and mechanisms come together they become something new, a new object with a new property: namely conscious experience. But what is this new object that has a phenomenal subjective point of view? It is not sufficient that the neuroscientific parts and mechanisms are present, LeDoux has to show that they together have transformed into a new part that can account for the new property. LeDoux does however not address this issue.

LeDoux cannot give a plausible account for how the conscious can emerge from the nonconscious ingredients in humans. What he knows is that these ingredients are necessary. He has investigated the effect of manipulating these ingredients and he knows that they must all be in place for the subject to be conscious of the stimulus, based on their verbal report. But he is not able to explain the relation between the conscious and the nonconscious parts. He hence does not have a plausible explanation of consciousness and therefore he cannot exclude consciousness in animals.

6.1.3.2 His explanation of behavior

LeDoux's explanation of animal behavior can be viewed as evidence that supports his theory. But can his explanation give a plausible account of behavior? One of the challenges with LeDoux's explanation of behavior is that he uses terminology at convenience. The part of his model (LeDoux, 2015, p. 45 Fig. 2.7) that he claims explains animal behavior refers to terms such as threat stimulus, threat detection, defense response control, defensive motivational state, defensive actions and reactions, brain areas, cognitive systems: attention, working memory etc. He argues that these are causally related to generate defensive behavior. He applies terminology from many different disciplines such as neuroscience, behavioral science, cognitive science and computational science. He refers to the molecular level, the cellular level, network level, the behavioral level as well as the mental level (for the human version), all in one model. Some of these play a proximate role whereas others are functions. It is not clear that these can be translated into causal factors that relate in the way he needs in his model. It is difficult to see whether or not the terms carry over between models and levels.

Let us take a look at learning. At the neuroscientific level learning involves synaptic modification. Learning is a concept and its extension is the instances of bodily movements that change after the encounter with a stimulus. It seems odd though to apply the term "learning" at the molecular and cellular level since there is no "learner" that plays a role at this level in the model. It is the individual that behaves, is rewarded or punished and learns, but LeDoux is using these anthropomorphic terms on cells and molecules. Molecules follow chemical and physical laws; synapses modify through cellular processes that manifest increasing or decreasing firing probability. These do not learn in the sense he is trying to explain. His explanation is restricted to description of molecular regularities. His attempt at reaching a purely neuroscientific objective explanation fails to capture the phenomenon he set out to explain. One cannot reduce a phenomenon to another explanatory level and still use the same vocabulary, at least not without making this translation explicit. LeDoux's project assumes that we already know what learning is; he does not give an account that allows us to construe learning in a more basic vocabulary.

LeDoux's model must be able to explain how animals can behaviorally adapt to changing environments: modify existing behavior (non-associative), learn to make associations between stimuli and connect this to behavior, and learn new behaviors, and it is not clear that his model can do this. He describes the processes in the synapses but he cannot relate it to how the animal can evaluate the behavioral success. He explains reinforcement in terms of the relation between the amount of stimuli present, motivational state and behavior. But this leaves open how the animal can "know" which stimuli are good and which are bad so that it can respond optimally, i.e. increase or decrease its behavior towards it. He does refer to a nonconscious value system in the amygdala that may be able to process value in the brain, but does not elaborate how. He has to relate the learning at the behavioral level with the synaptic modification. His explanation is not exhaustive in these regards.

6.1.4 Conclusion

LeDoux argues that animal behavior can be explained by the physical stance, by a neuroscientific model. Even if this is also the case for humans, he still appeals to the intentional stance because this is necessary to account for the storyteller in humans. He believes he has reliable evidence of this from verbal report and neurophysiological similarity to himself, and he can explain how consciousness emerges from nonconscious ingredients in humans. I have argued that LeDoux's theory of consciousness cannot account for how consciousness can emerge from these nonconscious ingredients. Therefore he cannot exclude animal consciousness. It is not clear that his explanation of learning can account for reinforcement and the relation between assessment of behavioral success at the behavioral level and modification of the synapses at the neuroscientific level. The terminology does not transfer between explanatory levels. His explanation of learning is thus not exhaustive, and cannot be used as evidence for this theory.

In the next section I will examine LeDoux's model of the relationship between stimuli, consciousness and behavior to see whether his model is plausible.

6.2 The relationship between stimulus, consciousness and behavior

The common model for the relationship between stimuli, consciousness and behavior is the following: A stimulus causes a conscious experience, which in turn causes behavior. For LeDoux, this is what happens in humans. In animals however, there is no conscious experience of any kind; stimuli nonconsciously cause physiological and behavioral responses. He uses the feeling fear as a case in point: fear is not necessary for eliciting defensive behavior in the encounter with a dangerous stimulus. He then generalizes this to be the case for all kinds of conscious experiences and all kinds of animal behavior. In this section I will identify and challenge two of the underlying assumptions in his model. Then I will see how this affects the interpretation of his examples. From this model I will develop a new marker of consciousness.

6.2.1 Stimuli, consciousness and behavior

LeDoux refers to cases that he believes illustrate that humans can respond physiologically, emotionally and behaviorally to stimuli processed nonconsciously. In these examples, presented previously, the subjects are prevented from consciously perceiving the stimuli due to blindsight, visual neglect, or subliminal stimulus presentation. Hence they are not able to verbally report the causal stimulus. He also presents examples of split-brain patients that express emotional reactions to stimuli that they cannot verbally report (Gazzaniga & LeDoux, 1978). The sequence of events in the examples is the following: the subjects are presented with stimuli and their response is recorded. LeDoux assumes that because the subjects do not verbally report the stimuli, they have not had conscious perceptions of the stimuli and therefore the process has been nonconscious. He thus concludes that conscious perception of stimuli is not necessary to induce physiological, emotional and behavioral responses in humans.

LeDoux's model of the relation between stimulus, consciousness and behavior makes two assumptions:
1. The causal part of a conscious experience is constituted by the properties of the stimulus.
→ If the properties of the stimulus are not perceived (verbally reported) then consciousness is not involved.
2. The causal role of a conscious experience comes after stimulus presentation and before behavior.

I will raise objections to these two assumptions.

6.2.1.1 Assumption 1: The causal content of consciousness

In the discussion on the relationship between stimuli, consciousness and behavior it is important to be clear about what exactly the conscious state concerns and what characterizes this content. Is it the US, the CS or the relation between these two (Åsli & Flaten, 2012)? Does a conscious state involve just the object, or does it also involve knowledge about the object? Furthermore, does the state also have to involve some kind of attitude towards either the object or the knowledge about this object? For example, if I perceive a red apple this can invoke many sensory perceptions in me: I see the shape, color, texture and movement, I feel the surface and I smell it. Even at this level it can be argued that there is aboutness involved because the information is *about* the object that is perceived. The perception of the apple will also invoke factual knowledge: I will experience specific factual knowledge, such as object recognition that it is an apple, and that its color is red, and general factual knowledge such as that apples are a fruit and grow on trees. It will also invoke episodic knowledge, for example I can remember that I have eaten apples before and I really like the taste. It may also invoke a feeling of desire or even fear – if the last time I ate an apple there was a worm inside it.

According to LeDoux, having a conscious experience of an apple involves all these experiences. Could it however be the case, that none of these requirements are necessary for a conscious experience to occur in the encounter with an apple? That is, I might not have to have knowledge of any categories or concepts or labels of any kind in order for the apple to invoke a conscious experience in me? I will argue that the basic causal content of a conscious experience is in fact not the stimulus, or the properties of the stimulus, but rather the behavior the animal is performing *in relation to the stimulus* and the assessment of this performance. Hence consciousness at the basic level is not about how the stimulus makes you feel, but how your behavior towards the stimulus makes you feel. This means that, strictly speaking, the subject, to be consciously encountering some object, need not know anything about the stimuli as such; the experiential content need not go beyond its own behavior in relation to the stimulus.

The common view is that the properties of the stimulus elicit specific neural circuits that cause conscious thoughts and feelings that in turn affect behavior. The empirical literature does not always specify what is meant by the subject being conscious, but it can usually be taken to mean that the subject is conscious of a specific stimulus. If one were to draw a thought bubble from the subject, the stimulus would figure as the content. The assumption is that the stimulus with its properties is assessed and that constitutes the source of the experience. For example, Kandel (2000, p. 1248) writes: "..an animal learns about the properties of a novel stimulus that it is harmless."

Different stimuli mean different things to different organism in that they affect them differently: what is toxic to one animal may be beneficial for another. Furthermore, some stimuli are necessary only in certain amounts. Salt is crucial for many species, but can be deadly in large doses. It can be argued that the dangerous aspect of the stimulus does not lie in the stimulus itself, but in the way that it can affect the animal. The only way an animal can affect its encounter with the external world is through its behavior. It has direct access, subjective knowledge about its own behavior.

Behavior, most basically, is considered bodily movement. Sensory cells are activated and the animal has direct access to these activations. But behavior may be more than outward bodily movement in the traditional sense of the term behavior. Nöe (2009) argues that for example seeing is a bodily activity in the sense that it involves moving the eyes, the head and the body. You have to put your sensory organs in the right position so that the brain can receive necessary information from the environment. You may have to squint your eyes, or focus your sight, in order for the information to enter the eyes. The goal is to see clearly and seeing clearly is rewarding. During seeing behavior, the eyes detect patterns in changes of photons. Similarly, hearing behavior involves changes in air pressure and smelling behavior involves changes in molecular concentrations in the air. The content of consciousness is the activity of perceiving the object, not the object in itself. On the view presented here, everything the animal can know about the external world is provided through assessment of behavior in relation to the external stimulus, not by assessment of the external stimulus itself. The stimulus affects what the experience feels like, but does not determine whether or not the animal is conscious. A consequence of this view is that lack of stimulus perception of a specific object preceding behavior does not imply that the subject is not conscious.

Some examples can be helpful to illustrate the advantage of being able to assess own behavior. Example 1: An automatic lawn mower can detect the boundary of a garden, marked by a cable, and this causes it to change direction. We do not believe that this involves consciousness. The machine has been programmed to change direction when it detects a cable, and the property it detects is electromagnetism. The machine does not have to identify the cable, or have any kind of knowledge, attitude or feeling towards the cable for adequate response to occur. The machine has been programmed to respond in a specific way to a specific stimulus. When we explain why the machine changes direction we say: "The cable caused it to change direction." The stimulus caused the response. For LeDoux, this is in principle how animal behavior can be explained. They are automata.

Example 2: An animal moving towards a wall for the first time blocking its path will first try to traverse the wall, and then change direction. It will experience its behavior towards the wall and this may involve pain if it keeps trying. Next time it reaches a wall, it has learned to associate cues and changes direction before it runs into it. When a human toddler sees an unfamiliar wall blocking her path, she may also try to traverse it. The next time she will change her direction. As she grows older the sight of the wall will come to contain more information, recognition and knowledge *about* the wall. We therefore assume that the stimulus itself (the wall) causes the intentional aspect in the subject, and that this is what affects her behavior.

The behavior of the lawn mower is to change direction in response to the boundary cable, and this has been programmed based on what is beneficial under current specific circumstances. Were however the environment to change, for example new objects that appear and block the path, then the machine would not be able to modify its behavior and improve on the next encounter. It would have to be programmed to handle every possible situation. There would be one specific behavior for every stimulus. The behavior is not flexible. Although much is programmed genetically in the neural system of animals (and humans), they are able to modify their behavior towards a stimulus from one encounter to the next.

In order to modify behavior towards stimuli and improve its interaction with the world, an animal needs to be able to evaluate or assess how successful a behavior is in the encounter with a specific stimulus. An explanation of behavior hence needs to be able to account for how a system can assess and modify its behavior in a way that improves survival and reproduction. On the view presented here, the phenomenal aspect *is* the assessment of behavior itself, what it feels like to perform a certain behavior in response to a stimulus.

Humans categorize specifics sets of movements into kinds of behaviors. These are labeled by words, often verbs. Examples of behavioral categories can be: run, investigate or drink. A category is often closely linked to what it considered the function of the behavior. Running may for example be categorized as “escape” behavior. But it can also be described as: “repetitive leg movement up and down pushing backwards.” Most behaviors involve an enormous number of movements. This whole chain of events is a continuous ongoing process and it can be difficult to define where each behavior starts and ends. For example "having a drink of water" encompasses the movements before and during the drinking itself: grabbing the glass and moving it towards the mouth, swallowing etc. This can be further broken down into smaller steps: drinking involves movements in the mouth, the throat etc.

Motivations and intentions may easily slip into these categories, and may confound which behavior is actually being reinforced. For example, "turning on the light" can be behaviorally described as "moving the finger towards the switch and pushing it up." That this has resulted in the light being turned on in the past – certainly affects the assessment of the behavioral steps as successful, but turning it on in the future is not what is driving the behavior. Basic behavioral units can be put together and the success of different sequences tested out through trial and error. In the end the system will be able to reinforce whole sequences of steps, which together comprise a behavior. What exactly counts as a basic unit and which sequences together constitute a behavior can be difficult to determine. How much can a behavior be modified until it counts as a new behavior is also difficult. Categorizing behavior of other species is even more challenging in terms grouping together the relevant behavioral steps.

Noë (2009) argues that consciousness is the result of the body's encounter with the world. He compares it to the way the musician can play music on an instrument: The body is an instrument and the environment affects it and this creates conscious experience. A brain and a nervous system is this not sufficient to create conscious experience. You need something to play on it. As he clearly states (Nöe, 2009, p. 165): "The brain is not, on its own, a source of experience or cognition. What gives the living animal's states their significance is the animal's dynamic engagement with the world around it."

The intuitive idea that a conscious state is about the stimulus rather than own behavior may be due to how the world manifests itself to humans. Human conscious experience is affected by the way we categorize, conceptualize and label the external world. Most likely this has advantages in that it allows for analysis and understanding of events in the world, which in turn makes us capable of making predictions about behavior of other humans, animals and plants and to communicate and share this knowledge. But it may also cause us to experience the world as if the properties we perceive are “out there.” On the assumption presented here, this is not the essential part of conscious experience. Consciousness is being able to feel your behavior in relation to stimuli and for this experience to affect the future performance of behavior towards these stimuli.

A neural system can categorize the world in the sense that the neural nets respond to grouped data and act on these, but it is not necessary that the system has any knowledge about what is perceives. For example, an animal can perceive another animal and respond adequately without having any factual or episodic knowledge about the animal. But it still has a conscious experience of its seeing behavior; it feels like something to be seeing the animal and its action towards it feels like something too. Hence language and thinking, what LeDoux consider to be requirements for consciousness, are not necessary for consciousness on the view presented here.

The reason that the stimulus has taken the leading role in explaining reinforcement may be due to what has been presented earlier, that there has been a shift in the literature on what constitutes the reinforcer: from subjective feelings to the stimulus itself. A reward is the ability of a stimulus to elicit approach responses, and reinforcement is the tendency for a stimulus to strengthen a response (White, 1989). In LeDoux’s animal model the stimulus acts as a nonconscious reinforcer of behavior through cellular and molecular events.

LeDoux requires that a conscious subject have explicit knowledge, about experience and about itself. I have argued earlier why conscious experience precedes labels and that labels are not necessary for conscious experience. The causal content of consciousness is the animal’s own behavior in relation to the stimuli it encounters.

In the next section I will challenge another assumption commonly made about the relationship between stimuli, consciousness and behavior, namely the timing component of the causal role of consciousness.

6.2.1.2 Assumption 2: The timing of the causal role of consciousness

It is common to assume that if consciousness is to play a causal role, then the conscious experience comes after stimulus presentation and before the performance of behavior. The idea is that the stimulus causes a conscious experience, which in turn affects behavior. This assumption is based on our intuitive feeling of agency: First I have a conscious perception of the stimulus, and this may in turn invoke knowledge and feelings towards the stimulus, and then this affects my action. The studies presented by LeDoux make this assumption: They present a stimulus, record behavior/physiological changes and then ask the subject if they could verbally report the stimulus that was presented. The assumption is the following:

Stimulus presentation → conscious experience → behavior

LeDoux's main focus is on the feeling fear and whether or not it is necessary to feel fear in order to perform defensive behaviors. In his model (LeDoux, 2015, fig. 2.7, p. 45) he illustrates his hypothesis: the dangerous stimulus can elicit defensive behavior directly, as well as the defensive motivational state – and even though this motivational state may in turn elicit feeling of fear in humans, feelings are not necessary to elicit the defensive behaviors. LeDoux concludes that conscious experience is not necessary to explain behavior.

In contrast to this, there are theories that do refer to consciousness when explaining behavior and I have presented these earlier. In these theories subjective feelings guide animal behavior and explain behavior in terms of the animals behaving to avoid pain and obtain pleasure. But on these views the reinforcer is the feeling associated with the *consequence* of the behavior. When the rat pulls a lever to obtain a food pellet, it is the food pellet that is doing the reinforcing work. Hence the feeling of pleasure, the conscious experience, follows after the behavior it is supposed to explain:

Stimulus presentation → behavior → conscious experience.

The reinforcement is hence explained in terms of conscious experience that follows after the behavior. This explanation is arguably functional and thus refers to something that occurs in the future. This is a "why" explanation according to Dennett (2015). We explain this behavior as: the rat pressed the lever so it could obtain the food pellet. This is a different sequence of events than the common view, but it is a different kind of explanation. We can also explain this in terms of the intentional stance, still in terms of preceding events: the rat had a belief that if it pulled the lever it would receive a pellet. This is how we would explain a similar situation with a human. I wish however to be able to describe the behavior in terms of proximate events at the behavioral level, not in terms of future events.

In the previous section I argued that the conscious content is not the stimulus itself, but rather what the behavior, in relation to the stimulus, feels like. This affects the future behavior in encounter with the stimulus and the sequence of events will look like this:

> Stimulus presentation → execution of behavior → conscious experience/assessment of behavior → modification of future behavior → stimulus presentation → execution of improved behavior...

At first glance, it may seem that focus on the conscious experience as reinforcer rather than the stimulus itself, is equivalent to the "why question" or the functional aspect. But this explanation differs from the ultimate explanation in that the conscious experience affects future behavior *after the next stimulus presentation*. What reinforces behavior is not the consequence of the behavior in terms of potential future conscious experience, but the conscious experience related to the *previous* behavior. The animal does not behave in order to obtain rewarding feelings, it acts based on what was experienced as rewarding the last time it behaved in this way in the encounter with the stimulus. This means that conscious experience of previous behavior is the reinforcer. Hence the previous performance becomes the cause rather than the future performance.

This learned material feeds into the causal chain of events that precedes execution of behavior, along with other processes that affect behavior. This is somewhat counter-intuitive for human thinking and feeling due to our tendency to interpret events as if we have an intention with our behavior, and hence this intention is commonly treated as a cause. Even though humans may believe that we behave to obtain rewarding feelings, and an animal may be motivated to obtain a reward, the causal role of the reinforcer is not from the future, but from the past. Given the complex nature of human conscious experiences, which involve mental time travel, it might be difficult for a human to determine if their conscious experience is before or after the behavior performed.

The stimulus causes a behavior, the animal experiences the success of the behavior and this experience affects the probability that the behavior is performed in the future. As discussed in the previous section, behaviors are large sets of movements. By continuously assessing the success of the movement, however small, the organism can test out the success of combinations. Consciousness hence becomes an ongoing process continuously assessing every little movement successively. This is in principle simple, but in reality extremely complex as behavioral sequences and assessment in themselves become stimuli that affect the subject. Hence at any time there is a myriad of successive stimuli being assessed making it practically impossible to disentangle.

I have argued that the causal content of conscious experience is performance of behavior in relation to stimulus. Conscious experience indirectly affects future performance of this behavior through learning. Based on this, I will now take a new look at LeDoux's examples and consider what conclusions they can support.

6.2.1.3 A new look at the examples

LeDoux has two groups of examples that he uses as evidence for his claims about the relation between consciousness and behavior. The first group of examples shows that humans can perform behavior and respond physiologically without conscious experience of the stimulus involved. Based on this he concludes that conscious experience is not necessary for behavior to be produced. The second group of examples demonstrates how and why humans are conscious. Consciousness in humans involves a storyteller: a verbal interpretation of own behavior to the self. This explicit interpretation is what constitutes the conscious experience. Hence without this verbal interpretation, and the necessary ingredients for this - such as language and the concept of a self - there can be no conscious experience.

I will now take a new look at these two groups of examples based on the two assumptions argued for in the previous section.

1.1.1.1.1 Group 1: Humans can perform behavior without being conscious of the causal stimulus

Based on the assumption that the causal content of consciousness is behavioral performance itself, and that sensory perception is assessment of visual behavior, the examples with blindsight, visual neglect and subliminal stimuli involve at least two behaviors: seeing and the behavioral response to the stimulus which can be either verbal or some (coaxed) physical task. It can be argued that even though the subjects report that they do not see the stimuli that cause their behavior they are still conscious in their seeing behavior. The terms "look" and "see" can illustrate this distinction: "I am looking but I can't see anything" versus "I am looking and I can see the object." The difference in these two statements is not whether you are conscious or not, in both cases you are performing seeing behavior and having phenomenal experiences; the difference is what you are perceiving during behavioral performance. This difference is what accounts for the difference in phenomenality: it feels different to see an object than to not see an object.

LeDoux may respond that this is just play with words (semantics): claiming that a person has two different perceptions, "a perception that includes object A" and "a perception which does not include object A", amounts to the same as "being conscious of A" and "not being conscious of A". But requiring that the animal have to consciously perceive the stimulus in order to be considered conscious misses the essence of what consciousness is about; it does not distinguish between being conscious and what the conscious state feels like. An animal is conscious of its behavior towards a stimulus. If the behavior towards the stimulus does not involve the stimulus, then future behavior towards this stimulus will not be improved. But the animal is still conscious. In the experiments, the human subjects do not report seeing the stimulus. Even though LeDoux is probably not claiming that these subjects are unconscious at the moment they miss seeing the stimulus, this is what he concludes about the animals. But whether or not the subject perceives the stimulus as part of the experience is not what determines if the subject is conscious or not.

The subjects in the experiments are performing seeing behavior in the sense that they are placing their eyes in the right direction and actively using them. But in these different experiments specific circumstances prevent them from having a phenomenal experience of the stimulus. But how then can these subjects still respond to the stimulus? They can, if one assumes that their behavior is elicited by nonconscious processes. And this in fact what LeDoux suggests. But this does not imply that animals do not need to see. Animals need to be able to assess behavioral success in relation to the stimulus and this is what they are doing through perceptions.

LeDoux also refers to situations that demand fast action. For example, in the encounter with a snake we act before we have time to see or feel the stimulus. Clearly the subject is not nonconscious during the event. But the fact that she can behave without having first having a conscious perception of the snake demonstrates that the brain can detect, process and execute behavior nonconsciously. The perception and feelings of fear are experienced afterwards. But LeDoux takes this to mean that they do not influence behavior. But based on the assumptions that conscious experience of behavior affects future behavior, his examples in fact support this.

What these examples demonstrate is that behavior can be elicited nonconsciously. They do not however demonstrate that the subjects are not having conscious experiences during behavioral performance and that conscious experiences do not play a causal role for future behavior. Here it is argued that quite the contrary: even though consciousness is not the elicitor of current behavior, consciousness is the *assessment* of current behavior and therefore affects the next instance of behavior. There are many events and processes that causally affect behavior, both external perceptions and internal bodily affairs such as energy levels, hormonal levels etc. It is safe to say that every behavior relies on a myriad of preceding events, such as described by LeDoux. One of the factors that affect behavior is what the organism has previously learned. Learning is hence an input to behavior; the learned material is likely to be just one of many factors that contribute to the causal chain involved in eliciting behavior. Panksepp (2011) argues similarly that feelings play an important role in reinforcing behavior, rather than executing or controlling behavior.

The relationship between stimulus, consciousness and behavior is thus the following: stimuli trigger behavior through nonconscious processes, and behavior in relation to stimuli is *evaluated* through conscious processes. There is no need for any kind of knowledge, categories, concepts, labels or self-reflection to elicit behavior, nor are these necessary for an animal to be able to assess its behavioral success. An animal only needs to have direct access to current experience of own behavior in relation to the stimulus. It does not need to be able to think about these experiences.

For humans it is common to associate conscious processes with processes that we can control. For example, we often assume that conscious learning implies conscious control of the behavior learned. But conscious experience does not necessarily imply conscious control. Hommel (2013) argues that consciousness plays *no* role in the control of action. If humans had conscious action control we should be able to report the reasons underlying our actions, but this is not the case (Wegner, 2002). This can be illustrated by the fact that humans are easily manipulated: We can be fooled into believing artificial limbs are part of our own body (Ehrsson et al., 2005) and even believe the actions of other people are our own (Nielsen, 1963). Conscious experience takes time to unfold and reach conscious awareness, every conscious moment lasting for about 300-500 ms (Libet, 2004). This is too slow for it to play a direct preceding role for the execution of behavior. This time factor alone is enough to suspect that much of our behavior is generated by nonconscious processes.

The idea that behavior is caused exclusively by nonconscious processes has been disputed. According to Stafford (2014) it is actually very difficult to demonstrate that a subject is unaware (Newell & Shanks, 2014). Furthermore a subject cannot always know for sure whether or not they were conscious of the stimulus; they may even misjudge their experience. Stafford (2014) believes that much of the support for nonconscious behavior execution comes from the fact that there has been a shift in the psychological literature on what is meant by nonconscious processes from "without awareness of the stimuli" to "without awareness of the influence of the stimuli", likely due to a definition put forward by Bargh (1992). The second definition requires that the subject can know how a stimulus influenced them. Given all the complex events that underlie behavior, this seems almost impossible. Hence studies based on this second definition will often conclude that the subject is nonconscious.

These criticisms of the idea that nonconscious processes are responsible for behavior are however based on the same assumptions that LeDoux makes, namely that what constitutes a conscious experience is the stimulus itself and it is this perception that causes behavior. What they challenge is whether or not a subject can always know and report whether or not they were conscious of the stimulus that assumedly preceded the behavior or even how it influenced them. But based on the assumptions argued for here, the stimuli are not processed consciously until behavior is performed; the behavior is elicited by nonconscious processing of the stimuli. The objections do however show that it is very difficult for humans to determine how and when their conscious experiences affect behavior.

In this section I have argued that the examples support the view that behavior is elicited by nonconscious stimuli. However, the examples do not demonstrate that the subjects are not conscious. In LeDoux's model the object of assessment in consciousness is the stimulus. On the view presented here the object of assessment is performance of behavior in relation to stimuli. This conscious assessment affects the future behavior towards the stimulus through nonconscious processes. Hence consciousness indirectly affects future behavior.

1.1.1.1.2 Group 2: Consciousness is an interpretation of behavior that involves a storyteller

According to LeDoux, the function of consciousness is to be a storyteller to the subject. He argues that conscious experience is an interpretation of the events and processes that occur inside and outside the subject. Language, with symbols, concepts and explicit labels, plays a necessary role. The function is to allow a subject to understand her own behavior and furthermore communicate it to other humans. Animals do not possess the necessary circuits or psychology for either self-representation or language hence they are not conscious.

There is evidence that humans generate explanations of own behavior based on how they actually behave, both from subjects with brain damage, but also from subjects with normal brains. I have previously presented LeDoux's examples of cases with split-brain patients (Gazzaniga and LeDoux, 1978). This indicates that the brain first elicits behavior, interprets it and then making up a plausible reason for it. LeDoux concludes that these examples show that emotional systems and detection of stimuli can operate separately. It can however be argued that actually these examples illustrate what I have argued for in the previous section. For example in the case of the woman who giggles when presented with nude pictures: the stimulus (nude pictures) nonconsciously elicits behavior (laughing), and this behavior is what the subject is experiencing or feeling as "funny." When asked about her laughing behavior the subject tries to find something that can be the cause of this funny feeling, and attributes it to the machine looking funny. Normally, the subject would perceive the stimulus in the conscious experience and be able to identify source of the emotion. What this example shows then is that the subject does not have to know anything about the stimulus in order to feel her own behavior. Having factual or episodic information about the stimulus is not necessary to have a conscious experience.

According to Damasio (2003, p. 67) there is close interplay between emotions, feelings and thoughts. He characterizes emotions as nonconscious physiological processes such as facial expressions, increase in heart rate or release of hormones. These processes can become conscious and give rise to feelings. He has an example of a woman, due to a brain procedure, received electrical stimulation in the brain. This induced spontaneous sad facial expression, followed by crying and then many sad thoughts. The brain was coming up with reasons to explain this crying. According to Cosmides & Tooby (2000), conscious emotional states can also elicit adequate cognitive programs.

All these examples seem to support LeDoux's idea that the storyteller in the subject is trying to interpret the feelings the subject is experiencing during behavioral performance. But they also demonstrate that the storyteller needs something to interpret in the first place: namely a conscious feeling. The examples demonstrate that the storyteller is sometimes wrong in its interpretation of the cause of behavior, but the feelings are not wrong. The conscious feeling thus seems to be able to operate without the storyteller. Based on this it can be argued that conscious feelings have to be present in order for the storyteller to be able to interpret own behavior. This is in contrast to LeDoux's theory that the interpretation made by the storyteller *is* the conscious experience.

It can be argued that what LeDoux is addressing is human thinking rather than consciousness itself. If consciousness at the basic level is assessment of own behavior through conscious experience, it is possible that by labeling these experiences through language humans can perform verbal assessment of behavior in an explicit sense rather than implicit. It is interesting to note that in his model the self interprets own behavior, rather than the stimuli; this goes well with the idea that assessment of behavior is the causal part of conscious experience.

The ability for story telling has many advantages. Language, with symbols and concepts allows for mental time travel and this has likely shaped our mind (Suddendorf, 2009). The ability to imagine arguably constitutes the fundament for human fictional reality and has likely provided huge survival advantages (Harari, 2011). On LeDoux's view the function of consciousness is to be a storyteller to the subject, which can cognitively inform the subject about the world and in this way improve behavior. On the view presented here, the function of consciousness is to assess behavioral success and in this way improve behavior. This view is however compatible with the idea of a human storyteller, but on this view the storyteller is not necessary for conscious experiences. Consciousness as a phenomenon precedes human verbal psychology. Conscious experience in humans is a chaotic myriad of perceptions, bodily sensations and thoughts, some more salient and enhanced others less so and maybe immediately forgotten. The cognitive interpretation, i.e. the storyteller, makes it challenging for humans to disentangle, identify, isolate and verbalize the more low level parts of their own experience.

6.2.2 A new marker of consciousness: learning

For an organism to modify and learn new behaviors, it has to be able to assess behavioral success. Consciousness as argued here, in its essence, is the ability to evaluate the performance of own behavior. Behavior is assessed thought sensory neurons. These are distributed over the body, some also internally. It is not a requirement that the experiential self has any explicit knowledge about the stimuli in the world. The model presented here suggests a new marker of consciousness: namely the ability of an organism to modify and learn new behavior in the encounter with stimuli in a changing environment.

I have argued that LeDoux's model of the relation between stimuli, consciousness and behavior is not plausible. Based on the new model I have presented, where behavior is elicited by nonconscious processes, what does this mean for behavior as a correlate of behavior? This means that behavior in itself is not enough. The organism has to be able to demonstrate learning: modification and learning of new behavior.

I have argued against LeDoux's requirement of explicit knowledge, however the issues he raises are indeed very interesting when it comes to the phenomenal question: How do other animals experience the world without explicit categories and labels? What does my dog see when we are walking in the forest, does she see trees, rocks and squirrels? What does she see when she is hormonally pseudopregnant and is collecting little toys under the bed as if they were puppies?

6.2.3 Conclusion

I have argued against two common underlying assumptions about the relationship between stimuli, consciousness and behavior. I propose that the causal content of consciousness is phenomenal experience of own behavior. This allows the animal to assess its behavioral success in relation to a stimulus. I have further argued that consciousness does not cause behavior directly. Behavior is elicited by nonconscious behavior, in agreement with LeDoux. Consciousness is the assessment of current behavior and plays a role for future performance of the behavior in relation to the stimulus. The current behavior is a result of the assessment of previous behavior in relation to the stimulus. This suggests a new marker of consciousness: the ability to modify and learn new behaviors in a changing environment. This model is compatible with LeDoux's idea of the storyteller, but in the model presented here conscious experience precedes the ability for explicit knowledge, hence this is not necessary component of conscious experience.

6.3 Consciousness and learning

An organism plays an interactive role in the world: it receives input in terms of stimuli, and produces output through own behavior that in turn become new stimuli that affect the organism. In a sense the environment is continuously programming the subject, and the subject is programming others both directly and indirectly through the stimuli they influence. The animal is continuously learning. Modification of behavior allows the animal to reinforce what works: behavioral elements can increase or decrease in frequency, and combinations of elements can be combined into new sequences. When behavior improves from one encounter to the next, the system learns (Simon, 1996).

In the previous sections I have argued, in agreement with LeDoux, that behavior can be elicited by nonconscious processes. But this does not imply that learning is nonconscious. Even though behavior and learning are closely related they are not the same processes, hence consciousness can be part of one and not the other. I have argued that LeDoux's explanation of learning and reinforcement cannot account for behavioral assessment in terms of nonconscious neuroscientific processes. The organism has direct access to conscious phenomenal experience of its own behavior, and this can allow the animal to experience behavioral success.

6.3.1 Synapses, computations and behavior

There are several levels at play. Execution of behavior relies on neuronal connectivity and activity. Learning involves change in this connectivity. At the neuroscientific level synaptic modification causes some circuits to be stronger linked than others, i.e. the synaptic weights increase and this increases the probability of signal transfer and hence increased probability of associated circuits to fire - circuits that underlie the future behavior. At the behavioral level a behavioral sequence either increases or decreases in frequency. In between the observed behavioral level and the neuroscientific level there is the computational level. Computational levels in the brain can be viewed as hierarchical abstractions (Ballard, 2015). A single neuron itself could be a conscious entity (Edwards, 2005) or the entity could be a result of networks of neurons. Many multicellular organisms have an integrated experiential center. Some may be only partly integrated, i.e. have several experiential center such has been argued may be the case for some animals (snake: Gärdenfors, 1995; for a discussion on frog an octopus: Godfrey-Smith, 2016). It is likely that the more body, in terms of size, limbs and movement options, the more there is to integrate in the experiential center, even though body size does not seem to be reflected in number of neurons (Herculano-Houzel, 2011). This requires higher abstract hierarchical levels.

6.3.2 The computational level

Some theories, like LeDoux's, assume that if a system is intelligent enough in the sense that it performs enough computations of the right kind such as self-representation, consciousness can emerge from this. But intelligence does not imply phenomenal consciousness. Dennett (2017) argues that a system can exhibit competence without comprehension, i.e. each part can be competent in their job without understanding the whole system. Both evolution and the Turing machine exhibit competence without comprehension. Similarly, Searle's Chinese room (1980) argues that a person or a machine can be competent at handling Chinese characters based on rules and produce meaningful combinations without understanding Chinese. Competence or intelligence is not sufficient to explain a mind.

In computer sciences, artificial systems have been designed to simulate neural nets in brains. These systems can also learn. It can be helpful to look at learning in artificial systems because these are realized solely by computations and this can illustrate the potential and limitations of the computational part of the system.

6.3.3 Learning in artificial systems

Artificial networks, also known as connectionist networks[4], are used to simulate the neural networks in the brain. These consist of nodes or units (neurons) and the connection weights between them (synapses). In these artificial systems behaviors or properties can emerge, which cannot be reduced to any particular unit but rather emerge from the ways in which the activations are spread throughout the network, not because they are built into the network. These networks naturally exhibit such phenomena as learning, generalization and pattern recognition.

There are three kinds of learning algorithms depending on the kind of feedback to the learner: Supervised, unsupervised and reinforcement learning. In supervised learning the system is told which output or target value it should produce for every input. It then gradually adjusts its connections weights between the hidden nodes to that the network improves and gradually learns to reach the desired output. This decreases the prediction error. The system learns to recognize (categorize) objects by learning to produce output patterns that represent the objects when presented with input patterns that represent the objects.

In unsupervised learning the learner does not receive this feedback from the environment. It infers functions that describe patterns or structures that are "hidden" in undefined data. The learner has to re-represent the input in a more efficient way, as clusters, categories or using a reduced set of dimensions, based on similarities and differences among the input patterns. The outputs are internal representations hence it does not result in direct change of behavior. But the representations can be used by

---

[4] Connectionist networks differ from classical symbol computational theory in their approach to understanding the brain and how it processes, represents and learns new information. Although both are computational theories of mind, they appeal to different architectural structures in the brain, and they make different assumptions and claims about which phenomena and properties play a role in the cognitive processes that underlie our cognition and behavior. The Turing machine operates with classical symbol computations. At the physical level however, the brain is a connectionist system of neural nets. It has been argued that the language-like concepts or symbols in these systems are postulated innate structures (Churchland, 2012).

other parts of the system that in turn will affect behavior. The connections that work will be reinforced, such as in Hebbian learning. In nature, unsupervised learning is much more common that supervised. Perceptual systems involve this kind of learning.

The third kind of learning is reinforcement learning. This is a kind of supervised learning in that the learner receives feedback based on its response. When it responds correctly then it receives information that this is an appropriate response, for example a positive value. When the response it wrong it gets feedback that the response inappropriate, and often how much, for example a negative value. However, it is not told what it should have done, which input/output pair is correct, like in supervised learning, but instead it is the behavior of the system that is reinforced.

In current Artificial Intelligence, nodes are reinforced by giving them feedback on what works and what doesn't work. This is also termed credit assignment function. This makes the system smart, even though each node is not smart. Previously, the best chess programs have been based on a combination of advanced search techniques, domain-specific adaptations and assessment functions developed by humans over many decades. However, the computer program AlphaZero recently taught itself to play chess in 24 hours and defeated world champion programs. It was not given any domain knowledge except for the game rules. It started out from random play, and learned through reinforcement learning during self-play and knowledge represented by deep convolutional neural networks. The designers claim that this shows that a general-purpose reinforcement algorithm can show super human performance without any previous knowledge (Silver et al., 2017).

The question is whether or not computations can be sufficient to explain the behavioral modification that we can observe in animals.

6.3.4 Learning in natural systems

I have earlier presented different kinds of learning in natural organisms. In nonassociative learning, the animal can modify its response in terms of frequency towards a reinforcing unconditioned stimulus. For example freeze in response to a predators painful teeth. This might stop the attack and it can escape. In classical conditioning, this response is expanded to be elicited by another, conditioned stimulus as well. The animal can now freeze to a new stimulus, which predicts pain, such as the predator's growl and this makes detection less likely. Instrumental conditioning allows the animal to modify components and compose new behaviors from these components based on what is rewarding or punishing. This allows for new behavior such as escaping by climbing a tree. All these kinds of learning in organisms involve reinforcement of behaviors. Even in nonassociative learning behavior is reinforced by the stimuli encountered, such as in the gill-withdrawal behavior of the sea slug (Kandel et al., 2000). The difference lies in how many stimuli can elicit the behavior, and how many different behaviors the animal can perform. Many simple reflexes can act as components of more complex reflexes, which together perform more complex responses such as locomotion (Zehr & Stein, 1999).

Reinforcement learning in machines seems to bear some similarity to learning at the behavioral level in organisms. Animals have been modeled in computers (Khamasse et al., 2005) and in synthetic models in the form of mechanical devices (Sporns et al., 2000) with artificial value systems and these "animals" exhibit reinforcement learning. Furthermore, artificial emotions in robots, such as fear, have proved advantageous for decision- making (Castro-González et al., 2013). There are however some differences. In artificial systems, someone has defined and preprogrammed the reward functions that describe how the system ought to behave. So the machine can develop new behaviors, but for these behaviors to be beneficial this requires that the environment does not change in ways that changes the rules in terms of what is rewarding and punishing. Furthermore, AI learns through hundreds of millions of samples whereas a human can learn from one sample.

Animals as well, need a reward function. But in nature the rules change. Animals have to adapt to ever-changing environments and therefore their reward function must be dynamic and be able to handle these changes. Some environmental factors will be present to a high degree and the organism can therefore be genetically prepared for these as "innate values" (Sporns et al., 2000). Other factors however will vary, both in between generations and during lifetime of the individual and therefore the values are acquired through experience. But what is this value system and how does it work?

6.3.5 Strange inversion, affordances and predictive coding

I have argued that sensory experiences are evaluative and that the organism has direct access during behavioral performance. This evaluation must come from the organism itself rather from the external world. This is compatible with what Dennett (2009) calls “strange inversion”: the idea that our experiences are projected from ourselves onto the external world. An animal’s perception of reality is created by the animal. The representations in its mind do not depict reality, but rather allow it to act successfully in the physical world (Sjølander, 1996). A lovely birdsong in my ear, can be the distress call of a mother looking for its chick.

Evolution has wired the organism’s brain so that for example, glucose triggers sweetness. The organism likes sweetness because it is good for it. This means that it is not the case that we like honey because it is sweet, but rather honey is sweet to us because we like it. The subject itself projects the sweetness, and all the other qualities, into the world. Through evolution our nervous systems have been designed to show us what is good or bad through reward and punishment. Dennett (2009) looks to Hume’s (1739) treatment of causation to explain this common human misunderstanding of mixing cause and effect: When A is followed by B this induces a feeling of expectation. However, we misinterpret this feeling and mistakenly believe that there is a perceptible property in the world that causes this impression. But this is an illusion, rather it is the other way around: the phenomenal experience is caused by the internal judgment of the feeling of expectation. We project the feeling out and attach it to the object, into the world and in this way misinterpret an inner reaction as an outer cause.

Dennett believes that sweetness, causation and colors are examples of what he labels “affordances” and that these yield a predictive action tendency. The term affordance was introduced by the psychologist Gibson (1975) and refers to what the environment offers the organism in terms of what is useful to it. Prediction is a forward model, which we then read backwards. For example, when we see a cup we have expectations about this cup and its affordances, for example that it can hold and carry liquid. In the brain there are Bayesian anticipation generators that cause us to project these perceptible properties to the cup. We are designed to perceive affordances that our essential for our survival and reproduction. What is experienced as familiar is constituted by the lack of prediction error.

Associative learning allows the animal to make predictions about what is to come. There is amounting evidence from computational neuroscience that the brain works as a hierarchical prediction machine (Grey, 1987; Clark, 2013). The brain generates predictions about the inputs it will receive. The difference between expected and actual reinforcement is called the prediction error. When inputs deviate from predicted values, the errors are corrected and the predictions are improved. The animal has to infer which stimuli give rise to its sensory signals. The brain is organized into levels, and the lower level systems receive input from the world. In higher-level systems a model of the casual structure of the world emerges and these systems attempt to predict the input to the lower-level systems. Errors cause the higher-level systems to adapt in order to reduce the discrepancy. It is these errors, deviations, which constitute conscious experience. When the actual input deviates from expectation, this leads to new learning. This predicts that if the reward is perfectly predicted, learning should not occur. This has been demonstrated in the blocking paradigm (Kamin, 1969).

Based on this it can be argued that the reward function in a natural system is realized through the deviation between expected and actual reinforcement. The deviation between expected sweetness and actual sweetness. This is conscious experience and conscious experience is learning. This is how the natural system can evaluate the success of a behavior. An implication of this is that not all behavioral performance is experienced consciously, only the experiences that deviate from the predicted.

Expectations can be genetically coded and inherited. Changes in the environment will select for the individuals with the best adapted reward predictions. For example, if a stimulus in the world that has previously been good for the organism becomes bad, the variants that projected this as less sweet will fare better. Research indicates that dopaminergic neurons in the basal ganglia in the midbrain may play a role: their phasic activity signals a discrepancy between the predicted reward and the currently experienced reward (Colombo, 2014). Hence dopamine is a possible currency (Ballard, 2015).

Reward functions likely play an important role in learning bodily control. Conscious experience can play an important role for skeletal muscle output (Morsella et al., 2016) Even though silicon computers can calculate must faster than humans, humans are superior with their musculo-skeletal systems to robots in the their ability to move our bodies (Ballard, 2015). Human control algorithms develop over years as we grow from babies. A human movement has many properties that play a role such as trajectory, speed and timing and in order to improve movement the brain needs to decide which properties should be modified and this is the credit assignment problem. Humans can quickly learn to perform the correct movements by the use of monetary reward without the target movement being explicitly defined (Dam et al., 2013).

Recent ideas on Attention Deficit Hyperactivity Disorder (ADHD) argue that this condition may involve problems related to the connection between the reward and the behavior (Mowinckel et al., 2015). The longer the delay between behavior and reward, the more difficulties in determining which behavior is the right one. The result is that the subject ends up generating and trying out many different behaviors.

An essential question is of course, why does the discrepancy between expected and actual reinforcement have to be conscious? As we have seen from artificial systems, that most likely are nonconscious, a preprogrammed reward function needs to be programmed by someone and it cannot handle unpredicted changes in the environment. In natural systems evolution is the programmer: genes code for the expected affordances and the system is dynamic is the sense that it adapts to the environment. Hence there must be a factor in the system, which can account for the assessment ensuring a dynamic system. Conscious experience of the deviation between expected and actual affordances may be a likely solution. The animal can consciously feel the success of behavior during learning.

Learning, manifested as modification of behavior and learning of new behaviors is then a positive marker of consciousness. This is a correlate that can be easily tested in many species. In fact, learning to avoid potentially painful stimuli, in addition to a reduction in adverse behavioral effects after administration of analgesics or painkillers, are two of the criteria that have been developed for pain perception in animals (Bateson, 1991).

A similar theory, has been proposed by Bronfman et al., (2016). They argue for unlimited associative learning (UAL) as the marker of a transition to sentience. This ensures that organisms can associate an unlimited amount of stimuli and generate an unlimited amount of reinforced actions. On this theory, it is not enough to demonstrate modified behavior in response to a stimulus, as in sensitization or habituation in the sea slug, because this is only modification of pre-existing behavior, there is not new behavior generated. They argue that their requirements are necessary in order for evolutionary sustainability. They argue that consciousness has emerged separately in arthropods, vertebrates and cephalopods. The requirement that the organism can learn an unlimited amount of behavior ensure that the animal can adapt to a changing environment.

6.3.6 Implications

There are several implications of the theory presented here. This is a functional requirement and it affects the neurophysiological requirements. Other species, at least those that differ largely from humans, may realize these functions by different structures. Learning as a correlate has implications for the distribution question. There are some organisms that appear not to learn, such as the cnidarian Box jellyfish (Jékely, 2011). This organism has diffused sensory and motor system, i.e. not a centralized brain. Such species must however be thoroughly explored before they are excluded. Habituation has been demonstrated organisms such as slime mould (Boisseau et al., 2016) and in plants (Gagliano et al., 2014). These organisms do not have neurons but have analog structures. Learning in plants is a recent field of inquiry, and it may be that more complex forms are demonstrated in the future. When it comes to Artificial intelligence the idea that future programming can produce a reward function with the necessary dynamics cannot be excluded, but it does not seem likely that they can come to acquire a phenomenal reward function.

6.3.7 Conclusion

Learning allows the animal to tune its existing behaviors and acquire new behaviors in a changing environment. Learning in a system requires a reward function to assess behavioral success. This has to be dynamic in the sense that it allows the system to adapt to a changing environment. Artificial systems have programmed reward functions but these systems cannot adapt successfully to changing rules. The animal experiences its own behavior; hence the reinforcement must come from within. Such reward functions in living systems can be genetically encoded and experienced through affordances. The brain can be viewed as a prediction machine that compares expected and actual reinforcement. Consciousness is the phenomenal experience of these the prediction errors, i.e. the discrepancy between expected and actual affordances. Consciousness is learning.

7 Concluding summary

The overall aim of this thesis was to identify some important tensions in the debate on animal consciousness, clarify positions and their relation to science. Joseph LeDoux has made some claims that challenge the idea that we can have knowledge about consciousness in other species. I have presented and discussed two claims. The first claim states that: Science cannot tell us about consciousness in animals. I argued that LeDoux's definition of consciousness is too narrow. Explicit knowledge about the experience and itself is not necessary for conscious experience. This narrow definition makes it challenging to demonstrate human consciousness. It reduces the reliability of the verbal report, and it arguably cannot demonstrate more than a behavioral substitute report by an animal. Science does not support the claim that there are significant differences between the human brain, and at least our closest primate relatives. I have concluded that if science, through verbal report and neural correlates of consciousness, can tell us about consciousness in humans, then it call tell us about consciousness in animals. Furthermore, several of LeDoux statements are inconsistent.

The second claim: I can explain animal behavior by neuroscience. From this LeDoux concludes that animals are not conscious. I briefly discussed aspects of explanations and I argued that his theory on consciousness does not succeed in explaining how consciousness can emerge from nonconscious ingredients in humans. He thus cannot exclude consciousness in animals. I examined his explanation of behavior and concluded that his explanation of learning and reinforcement cannot give a plausible account of how an animal can modify and learn new behaviors in a changing environment, in terms of cellular and molecular neuroscientific processes.

I then looked at his model of the relationship between stimuli, consciousness and behavior. I identified two underlying assumptions that I argued against. I proposed a model where the causal content of consciousness is the conscious experience of performance of behavior in relation to the stimulus, rather then the stimulus itself. Conscious experience does not directly cause behavior but the assessment of behavioral success modifies the next instance of this behavior in the relation to the stimulus. In this way consciousness indirectly affects future behavior. Behavior is caused by nonconscious processes, in agreement with LeDoux, and this model is also compatible with the idea of a human storyteller. But in this model the storyteller is an interpreter of conscious experience of behavior, and is not necessary for conscious experience, as in LeDoux's account.

Learning allows the animal to tune its existing behaviors and acquire new behaviors in a changing environment. Learning in a system requires a reward function to assess behavioral success. This has to be dynamic so that the system can adapt to a changing environment. Artificial systems have programmed reward functions but these cannot adapt to changing rules. An animal experiences its own behavior and the reinforcement must come from within. Such reward functions in living systems can be genetically encoded and experienced through affordances. The brain can be viewed as a prediction machine that compares expected and actual reinforcement. Consciousness is the phenomenal experience of these the prediction errors, i.e. the discrepancy between expected and actual affordances. Consciousness is learning.

I propose a new positive marker of consciousness: the ability to modify and learn new behaviors in a changing environment. This generates testable predictions and is a promising direction for the future studies of correlates of consciousness.

8 References

Akrami, A., Kopec, C.D., Diamond, M.E. and Brody, C.D. (2018). Posterior parietal cortex represents sensory history and mediates its effects on behaviour. *Nature*, Vol. 554, pp. 368-372.

Akins, K. (1993). What is it like to be boring and myopic? In: *Dennett and His Critics. Demystifying Mind,* B. Dahlbok (ed.), Blackwell, 247 pps.

Allen, C. & Bekoff, M. (1997). *Species of Mind*. MIT Press. 209 pps.

Allen, C. & Trestman, M. (2014). Animal Consciousness. *The Stanford Encyclopedia of philosophy* (Summer 2014 Edition). https://plato.stanford.edu/archives/sum2014/entries/consciousness-animal/

Allen, K.A., Coetzee, J.F., Edwards-Callaway, L.N., Glynn, H., Dockweiler, J., KuKanich, B. Lin, H., Wang, C., Fraccaro, E., Jones, M. and Bergamasco, L. (2013). The effect of timing of oral meloxicam administration on physiological responses in calves after cautery dehorning with local anesthesia. *J. Dairy Sci.,* 96, pp. 5194-5205.

Andrews, N., Legg, E., Lisak, D., Issop, Y, Richardson, D., Harper, S., Pheby, T. Huang, W., Burgess, G., Machin, I. and Rice, A.S.C. (2012). Spontaneous burrowing behavior in the rat is reduced by peripheral nerve injury or inflammation associated pain. *Eur. J. Pain*, 16, pp. 485-495.

Antonov, I., Antonova, I., Kandel, E.R. and Hawkins, R.D. (2003). Activity-dependent presynaptic facilitation and Hebbian LTP are both required and interact during classical conditioning in Aplysia. *Neuron*, 37, 1, pp. 135-147.

Aristotle. Four causes: *Physics* II 3 and *Metaphysics* V 2.

Azevedo, F.A., Carvalho, L.R., Grinberg, L.T., Farfel, J.M., Ferretti, R.E., Leite, R.E., Jacob Filho, W., Lent, R., and Herculano-Houzel, S. (2009). Equal number of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *The Journal of Comparative Neurology,* 513, 5, pps 532-541.

Baars, B.J. (1997). In the theatre of consciousness. Global Workspace Theory, a rigorous scientific theory of consciousness. *Journal of Consciousness Studies*, Vol. 4, No. 4, pp. 292-309.

Balduzzi, D. & Tononi, G. (2009). Qualia: The Geometry of Integrated Information. *PLoS Computational Biology*, Vol. 5, Issue 8, pp. 1- 24.

Ballard, D.H. (2015). *Brain Computation as Hierarchical Abstraction*. The MIT Press. 440 pps.

Balter, M. (2009). On the Origin of Art and Symbolism. *Science*, Vol. 323, pps. 709-711.

Bargh, J.A. (1992). Why subliminality does not matter to social psychology: awareness of the stimulus vs awareness of its effects. In: *Perception Without Awareness: Cognitive, Clinical and Social Perspectives*. R. Bornstein & T. Pittman (eds.), New York, NY, pp. 236-255.

Barron, A.B. & Klein, C. (2015). What insects can tell us about the origins of consciousness. *PNAS*, Vol. 113, No., 18, pp. 4900-4908.

Bateson, P. (1991). Assessment of pain in animals. *Animal Behaviour*, 42, pp. 827-839.

Bechara, A. et al., (1995). Double Dissociation of Conditioning and Declarative Knowledge Relative to the Amygdala and Hippocampus in Humans. *Science,* 269, pp. 1115-1118.

Bermudez, J.L. (2003). *Thinking without words*. Oxford University Press. 225 pps.

Berry, D.C. & Dienes, Z. (1993). Towards a characterization of implicit learning. In: *Implicit Learning: Theoretical and Empirical Issues*. Dianne C. Berry & Zoltán Dienes (eds.), Lawrence Erlbaum, pp. 1-18.

Berwick, R.C., Okanoya, K., Beckers, G.J.L. and Bolhuis, J.J. (2011). Songs to syntax: the linguistics of birdsong. *Trends in Cognitive Sciences*, Vol. 15, No. 3, pps.113-121.

Blackmore, S. (2016). Delusions of consciousness. *Journal of Consciousness Studies*, 23, pp. 52-64.

Block, N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavioral and Brain Sciences*, 30, pp. 481-548.

Block, N. (2011). Perceptual Consciousness overflows cognitive access. *Trends in Cognitive Sciences*, 15, 12, pp. 567-575.

Boisseau, R.P., Vogel, D. and Dussutour, A. (2016). Habituation in non-neural organisms: evidence from slime moulds. *Proc. R. Soc. B.*, 283, pp. 1-7.

Bolhuis, J.J. & Wynne, C.D.L. (2009). Can evolution explain how minds work? *Nature*, Vol. 458, 16, pp. 832-833.

Bronfman, Z.Z., Ginsburg, S. and Jablonka, E. (2016). The Transition to Minimal Consciousness through the Evolution of Associative Learning. *Frontiers in Psychology*, Vol. 7, Article 1954, pp. 1-16.

Broom, D.M. (1998). Welfare, Stress, and the Evolution of Feelings. *Advances in the Study of Behavior*, Vol. 27, pp. 371-403.

Broom, D.M. (2010). Cognitive ability and awareness in domestic animals and decision about obligations to animals. *Applied Animal Behavior Science*, 126, pp. 1-11.

Bshary, R., Hohner, A., Ait-el-Djoudi and Fricke, H. (2006). Interspecific Communicative and Coordinated Hunting between Groupers and Giant Moray Eels in the Red Sea. *PLOS Biology*, Vo. 4, Issue 12, pp. 2393-2398.

Call, J. & Tomasello, M. (2008). Does the chimpanzee have a theory mind? 30 years later. *Trends in Cognitive Sciences*, Vol. 12, No. 5, pp. 187-192.

The Cambridge Declaration on Consciousness (2012). http://fcmconference.org/img/CambridgeDeclarationOnConsciousness.pdf

Carey, S. (2009). *The Origin of Concepts*. Oxford University Press, 598 pps.

Castro-González, Á., Malfaz, M. and Salichs, M.A. (2013). An autonomous social robot in fear. *IEEE Transactions on Autonomous Mental Development*, Vol. 5, Issue 2, pp. 135-151.

Cervero, F. (2012). *Understanding Pain*. MIT Press, 172 pps.

Chalmers, D.J. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press, 414 pps.

Chalmers, D. J. (2010). *The Character of Consciousness*. Oxford University Press, 596 pps.

Chalmers, D. (2011). Verbal Disputes. *Philosophical Review*, 120, No. 4, pp. 515-566.

Chalmers, D. (2012). A Computational Foundation for the Study of Cognition. http://consc.net/papers/computation.html

Chittka, L. & Niven, J. (2009). Are bigger brains better? *Curr. Biol.,* 19, 21, pp. 995-1008.

Christiansen, M.H. and Kirby, S. (2003). Language evolution: consensus and controversies. *Trends in Cognitive Sciences*, Vol. 7, No., 7, pp. 300-307.

Churchland, P.M. (2012). *Plato's Camera. How the Physical Brain Captures a Landscape of Abstract Universals*. MIT press. 289 pps.

Clark, A. (2013). Whatever next? Predictive brains, situated agents and the future of cognitive science*. Behavioral and Brain Sciences*, 36, pp. 181-253.

Clayton, N., Emery, N., and Dickinson, A. (2006). The rationality of animal memory: Complex caching strategies of western scrub jays. In: *Rational Animals*. Eds: M. Nuuds and S. Hurley, Oxford University Press, pps. 197-216.

Cleeremans, A. & Jiménes, L. (2002). Implicit Learning and Consciousness: A Graded, Dynamic Perspective. In: *Implicit Learning and Consciousness: An Empirical,* Psychology Press. Robert M. French & Axel Cleeremans (eds).

Cleeremans, A. (2011). The radical plasticity thesis: how the brain learns to be conscious. *Frontiers in Psychology*. Vol. 2, No. 86, pp. 1-12.

Colombo, M (2014). Deep and beautiful. The reward prediction error hypothesis of dopamine. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*. Vol. 45, pp. 57-67.

Colpaert, F.C:, Tarayre, J.P., Alliaga, M., Bruins Slot, L.A., Attal, N. and Koek, W. (2001). Opiateself-administration as a measure of chronic nociceptive pain in arthritic rats. *Pain*, 91, pp. 33-45.

Cosmides, L. and Tooby, J. (2000). Ch. 7: Evolutionary Psychology and the Emotions. In: Lewis, M. and Haviland-Jones, J.M. (2000). *The Handbook of Emotions*, 2nd edition (91-115). The Guilford Press.

Crane, T. (1998). Intentionality as the Mark of the Mental. *Philosophy*, July 15, Issue supplement 43, pp. 229-251.

Craver, C.F. & Darden, L. (2013). *In Search of Mechanisms. Discoveries across the Life Sciences.* The University of Chicago Press, 228 pps.

Dam, G., Kording, K. and Wei, K. (2013). Credit Assignment during Movement Reinforcement Learning. *PLoS ONE*, Vol. 8, Issue 2, pp. 1-8.

Damasio, A. (1999). Ch. 2: Emotion and Feeling. In: *The Feeling of What Happens*, 35-81. London: Vintage Books.

Damasio, A. (2003). *Looking for Spinoza. Joy, Sorrow and the Feeling Brain*. Vintage Books. London, 339 pps.

Darwin, C. (1872). *The Expression of the Emotions in Man and Animal*, London, Fontana press.

Dawkins, M.S. (2008). The Science of Animal Suffering. *Ethology,* 114, 99. 937-945.

Deacon, T. (1997). *The Symbolic Species: The Co-evolution of Language and the Brain*. W.W. Norton & Company Ltd, 527 pps.

Dehaene, S. & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition*, 79 (1-2), pp. 1-37.

Dennett, D. (1991). *Consciousness Explained*. Penguin Books (1993), 511 pps.

Dennett, D. (1995). The path not taken. Commentary/ Block: Confusion about consciousness. *Behavioral and Brain Sciences*, 18, 2, pp. 252-253

Dennett, D. (2009). Darwin's 'strange inversion of reasoning'. *PNAS*, Vol. 106, Suppl. 1, pp. 10061-10065.

Dennett, D. (2017). *From Bacteria to Bach and Back. The Evolution of Minds*. Allen Lane, Penguin Books. 476 pps.

deWaal, F.B.M. & Ferrari, P.F. (2010). Towards a bottom-up perspective on animals and human cognition. *Trends in Cognitive Sciences,* Vo. 15, No. 5, pp. 201-207.

Eccles, J.C. (1965). *The Brain and the Unity of Conscious Experience*, The 19th Arthur Stanley Eddington Memorial Lecture, Cambridge, England: Cambridge University Press.

Edwards, J.C.W. (2005). Is Consciousness Only a Property of individual Cells? *Journal of Consciousness Studies*, 12, 4-5, pp. 60-76.

Ehrsson, H.H., Holmes, N.P. and Passingham, R.E. (2005). Touching a rubber hand: feeling a body ownership is associated with activity in multisensory brain areas. *J. Neursci*., 25, 45 pp. 10564-10573.

Elwood, R.W. (2011). Pain and Suffering in Invertebrates? *ILAR* (Institute for laboratory research) J., Vol. 52, No. 2, pp. 175-184.

Erber, R., Wulf, M., Becker-Birck, M., Kaps, S., Aurich, J.E., Möstl, E. and Aurich, C. (2012). Physiological and behavioural responses of young horses to hot iron branding and microchip implantation. *The Veterinary Journal*, 191, pp. 171-175.

Evans-Roberts, C.E.Y. & Turnbull, O.H. (2011). Remembering Relationships: Preserved Emotion-Based Learning in Alzheimer's Disease. *Experimental Aging Research*, 37, pp. 1-16.

Fitzgerald, M. (2012). The Biological Basis of Pain in Infants and Children. In: *Pain 2012*, Refresher Courses 14th World Congress on Pain, pp. 391-398.

Fuster, J.M. (2003). *Cortex and Mind. Unifying Cognition*. Oxford University Press. 294 pp.

Gagliano, M., Renton, M., Depczynski, M. and Mancuso, S. (2014). Experience teaches plants to learn faster and forget slower in environments where it matters. *Oecologia*, 175, pp. 63-72.

Gallistel, R. (2008). Learning and Representation. In: *Learning Theory and Behavior*. Ed. R. Menzel. Vol. 1 of Learning and Memory: A Comprehensive Reference, Ed. J. Byrne, pp. 227-242. Oxford: Elsevier.

Gallistel, C.R. (2011). Prelinguistic thought. *Language, Learning and Development*, 7, pp. 253-262.

Gärdenfors, P. (1995). Cued and Detached Representations in Animal Cognition. *Behavioural Processes*, Vol. 35, Issues 1-3, pp. 263-273.

Gardner, R.A. & Gardner, B.T. (1969). Teaching Sign Language to a Chimpanzee. *Science, New Series,* Vol. 165, No. 3894, pp. 663-672.

Gazzaniga, M.S. and LeDoux, J.E. (1978). *The Integrated Mind*. Plenum Press. New York and London. 168 pp.

Gentle, M.J. (2011). Pain issues in poultry. *Applied Animal Behaviour Science*, 135, pp. 252-258.

Gibson, J.J. (1975). Affordances and Behavior. In: *Reasons for Realism: Selected Essays of James J. Gibson*, E.S. Reed & R. Jones (eds.), Lawrence Erlbaum, Hillsdale, NJ, 1. ed, pp. 410-411.

Godfrey-Smith, P. (2016). *Other Minds. The Octopus and the Evolution of Intelligent Life*. William Collins, 255 pps.

Graham, K.E., Hobaiter, C., Ounsley, J., Furuichi, T. and Byrne, R.W. (2018). Bonobo and chimpanzee gestures overlap extensively in meaning. *PLoS Biol*., 16, 02, pp. 1-18.

Grandin, T. (2009). *Thinking in Pictures*. Bloomsbury Publishing, pps. 304.

Grey, J.A. (1987). *The Psychology of Fear and Stress*. 2. Edition. Cambridge University Press.

Hameroff, S. & Penrose, R. (2014). Consciousness in the universe. A review of the 'Orch OR' theory. *Physics of Life Reviews*, 11, pp. 39-78.

Hamilton, W. (1856). *Discussions on philosophy and literature, education and university reform.* New York, Harper&Brothers.

Harari, Y.N. (2011). *A Brief History of Human Kind.* Penguin Random House, 498 pps.

Harnad, S. (1987). *Categorical Perception: The groundwork of cognition*. New York, Cambridge University Press.

Harnad, 2016. Animal sentience: The other-minds problem. *Animal Sentience*, No. 1, 1-11.

Heider, F. & Simmel, M. (1944). An experimental study of apparent behavior. *American Journal of Psychology*, 57, pp. 243-259.

Herculano-Houzel, S. (2011). Brains matter, bodies maybe not: the case for examining neuron numbers irrespective of body size. *Ann. N.Y. Adac. Sci*., 1225, pp. 191-199.

Hommel, B. (2013). Dancing in the dark: no role for consciousness in action control. *Frontiers in Psychology,* Vol. 4, 380, pp. 1-3.

Hume, D. (1739). *A Treatise of Human Nature.*

Jékely, G. (2011). Origin and early evolution of neural circuits for the control of ciliary locomotion. *Proc. R. Soc. B.*, 278, pp.914-922.

Johnston, V.S. (1999). *Why we feel. The Science of Human Emotions*. Perseus Publishing. 210 pps.

Kahneman, D. (2013). *Thinking, Fast and Slow*. Ferrar, Straus and Giroux. 499 pps.

Kamin, L.J. (1969). Predictability, surprise, attention and conditioning. In: *Punishment and Aversive Behavior*, B.A. Campbell and R.M. Church, New York, NY, Appletom Century Crofts, pp. 279-296.

Kandel, E.R., Schwartz, J.H. and Jessell, T.M. (2000). *Principles of Neural Science*. McGraw-Hill, 1414 pps.

Kandel, E.R. (2017). Personal communication on e-mail.

Karin-D'Arcy, M.R. (2005). The Modern Role of Morgan's Canon in Comparative Psychology. *International Journal of Comparative Psychology*, 18, pp. 179-201.

Khamassi, M., Lachèze, L., Girard, B., Berthoz, A. and Guillot, A. (2005). Actor-Critic Models of Reinforcement Learning in the Basal Ganglia: From Natural to Artificial Rats. *Adaptive Behavior*, Vol. 13(2), pp. 131-148.

Lamme, V.A.F. (2003). Why visual attention and awareness are different. *TRENDS in Cognitive Sciences*, Vol. 7, No. 1, pp. 12-18.

Langer, E.J. (1989). *Mindfulness.* 25th Anniversary Edition, 2014, Da Capo Press, 246 pps.

LeDoux, J. (1998). *The Emotional Brain*, Phoenix, Orion Books, 384 pps.

LeDoux, J. (2011). Templeton Lecture. University of Sydney, https://www.youtube.com/watch?v=8yxgPFXWLJA, time: 15.09-15.28.

LeDoux, J. (2012). Rethinking the emotional brain. *Neuron*, 73, pp. 653-676.

LeDoux, J. (2014). Coming to terms with fear. *PNAS*, Vol. 111, No., 8, pp. 2871-2878.

LeDoux, J. (2015). *Anxious. Using the Brain to Understand and Treat Fear and Anxiety*. Viking, 466 pps.

LeDoux, J. and Brown, R. (2017). A higher-order theory of emotional consciousness. *PNAS*, www.pnas.org/cgi/doi/10.1073/pnas.1619316114

Lewis, D. (1986). *On the Plurality of Worlds*, Oxford, Blackwell.

Libet, B (2004). *Mind Time: The Temporal Factor in Consciousness*. Cambridge MA, Harvard University Press.

Low, P., Panksepp, J., Reiss, D., Edelman, D., Van Swinderen and B., Koch, C. (2012). The Cambridge Declaration on Consciousness. In: Francis Crick Memorial Conference, University of Cambridge, England.

Lycan, W.G. (2001). A simple argument for a higher-order representation theory of consciousness. *Analysis*, 61,1 pps. 3-4.

Malafoglia, V., Bryant, B., Raffaelil, W., Giordano, A. and Bellipanni, G. (2013). The Zebrafish as a Model for Nociception Studies. *J. Cell. Physiol*., 228, pp. 1956-1966.

Manson, N. (2000). State consciousness and creature consciousness: a real distinction. *Philosophical Psychology*, Vol. 13, No. 3, pp. 405-410.

Mercado III, E. (2016). Commentary: Interpretations without justification: a general argument against Morgan's Canon. *Frontiers in Psychology*, Vol. 7, Article 452, pp. 1-3.

Merker, B. (2007). Consciousness without a cerebral cortex: A challenge for neuroscience and medicine. *Behavioral and Brain Sciences*, 30, pp. 63-81.

Merker, B. (2017). Oral comment at the conference: Animal Consciousness. NYU, November 2017. Organized by Ned Block and David Chalmers.

Miller, A.M., Kitson, G.L., Skalkoyannis, B., Flecknell, P.A. and Leach, M.C. (2016). Using the mouse grimace scale and behaviour to assess pain in CBA mice following vasectomy. *Appl. Anim. Behav. Sci*., 181: pp. 160-165.

Millsopp, S. & Laming, P. (2008). Trade-offs between feeding and shock avoidance in goldfish (carassius auratus). *Applied Animal Behaviour Science*, 113, pp. 247-254.

Milner, D. & Goodale, M. (2006). *The Visual Brain in Action*, Oxford University Press.

Morgan, C.L. (1903). *Introduction to comparative psychology*. London, Walter Scott Publishing.

Morrison, J.H. & Salzman, C.D. (2010). Re-Valuing the Amygdala. *Current Opinion in Neurobiology*, 20, pp. 221-230.

Morris, J.S., Öhman, A. and Dolan, R.J. (1998). Conscious and Unconscious Emotional Learning in the Human Amygdala. *Nature*, 393, pp. 467-470.

Morsella, E., Godwin, C.A., Jantz, T.K., Krieger, S.C. and Gazzaley, A. (2016). Homing in on consciousness in the nervous system: An action-based synthesis. *Behavioral and Brain Sciences,* 39, pp. 1-70.

Mowinckel, A.M., Pedersen, M.L., Eilertsen, E. and Biele, G. (2015). A meta-analysis of decision-making and attention in adults with ADHD*. J. Atten. Disord*., 19, 5, pp. 355-367.

Mowrer; O.H. & Lamoreaux, R.R. (1946). Fear as an Intervening Variable in Avoidance Conditioning. *Journal of Comparative Psychology*, 39, pp. 29-50.

Musen, G. & Squire, L.R. (1991). Normal acquisition of novel verbal information in amnesia. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 17, pp. 1095-1104.

Nagel, T. (1974). What Is It Like to Be a Bat? *The Philosophical Review*, Vo. 83, No. 4, pp. 435-450.

Nagel, T. (1986). *The View From Nowhere*. Oxford University Press. 244 pps.

Newell, B.R. & Shanks, D.R. (2014). Unconscious influences on decision-making: a critical review. *Behav. Brain Sci.*, 37, pp. 1-19.

Nielsen, T. (1963). Volition: a new experimental approach. *Scand. J Psyshol.*, 4, pp. 225-230.

Noë, Alva (2009). *Out of Our Heads. Why You are Not Your Brain, and Other Lessons from the Biology of Consciousness*. Hill and Wang. 214 pps.

Packard, M.G., Hirsh, R. and White, N.M. (1989). Differential effects of fornix and caudate nucleus lesions on two radial maze tasks: Evidence for multiple memory systems. *Journal of Neuroscience*, 9, pp. 1465-1472.

Panksepp, J. (2005). Affective consciousness: Core emotional feelings in animals and humans. *Consciousness and Cognition*, 14, pp. 30 – 80.

Panksepp, J. (2011). The basic emotional circuits of mammalian brains: Do animals have affective lives? *Neuroscience and Biobehavioral Reviews*, 35, pp. 1791-1804.

Panksepp, J. (2015). Rewarding and Punishing Properties of Deep Brain Stimulation: The Most Promising Entry Points for Constitutive Studies of Affective Experiences in Other Animals…With Profound Psychiatric Implications for Human Consciousness and Psychiatric Therapeutics. *Psychology of Consciousness: Theory, Research and Practice*, Vol. 2, No. 1, pp. 24-29.

Panksepp, J. & Biven, L. (2012). *The Archaeology of Mind*. W.W. Norton & Company, 562 pps.

Parvizi, J. (2009). Corticocentric myopia: old bias in new cognitive sciences. *Trends in Cognitive Sciences,* Vol. 13, No.8, pp. 354-359.

Penn, D.C., Holyoak, K.J. and Povinelli, D.J. (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences*, 31, pp. 109-178.

Pepperberg, I.M. (2012). Symbolic Communication in the Grey Parrot. *The Oxford Handbook of Comparative Evolutionary Psychology*. Edited by Todd K. Shackelford and Jennifer Vonk.

Pererira, A.Jr., Edwards, J.C.W., Lehmann, D., Nunn, C., Trehub, A. and Velmans, Max (2010). Understanding Consciousness. A Collaborative Attempt to Elucidate Contemporary Theories. *Journal of Consciousness Studies*, 17, No. 5-6, pp. 213-219.

Pessiglione, M., Petrovic, P., Daunizeau, J., Palminteri, S., Dolan, R.J. and Frith, C. (2008). Subliminal Instrumental Conditioning Demonstrated in the Human Brain, *Neuron*, 59, pp. 561-567.

Rees, G., Kreiman, G. and Koch, C. (2002). Neural Correlates of Consciousness in Humans. *Nature Reviews. Neuroscience*. Vol. 3, pp. 261-270.

Rose, J.D., Arlinghaus, R., Cooke, S.J., Diggles, B.K., Sawynok, W., Stevens, E.D. and Wynne, C.D.L. (2012). Can fish really feel pain? *Fish and Fisheries*. Blackwell Publishing, pp. 1-37.

Rosenthal, D. (1993). Thinking that one thinks. In: *Consciousness*, ed. M. Davies and G. Humphreys, pp. 197-223. Oxford, Blackwell.

Rosenthal, D. M. (2005). *Consciousness and Mind*. Oxford University Press, pps. 386.

Sardi, S., Vardi, R., Sheinin, A., Goldental, A. and Kanter, I. (2017). New Types of Experiments Reveal that Neuron Functions as Multiple Independent Threshold Units. *Scientific Reports*, 7: 18036, DOI:10.1038/s41598-017-18363-1

Schaller, S. (1991). *A Man Without Words*. University of California Press, 210 pps.

Schon, K., Parker, A. and Woods, CG. (2018). *Congenital Insensitivity to Pain Overview*. In: Adam MP, Ardinger HH, Pagon RA, et al., editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2018. Available from: https://www.ncbi.nlm.nih.gov/books/NBK481553/

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

Searle, J. (2000). Consciousness. *Annual review of neuroscience*, Vol. 23, pp. 557-578.

Shulman, R.G. (2013). *Brain Imaging: What it Can (and Cannot) Tell us About Consciousness*. Oxford University Press, 173 pps.

Siclari, F., Baird, B., Perogamvros, L., Bernardi, G., LaRocque, J.J., Riednes, B., Boly, M., Postle, B.R. and Tononi, G. (2017). The neural correlates of dreaming. *Nat. Neurosci.,* 20, 6, pp. 872-878.

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., Lillicrap, T., Simonyan, K. and Hassabis, D. (2017). Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm.

Simon, H.A. (1996). *The Sciences of the Artificial*. 3. edition. MIT press. 231 pps.

Sjølander, S. (1996). On the Evolution of Reality – Some Biological Prerequisites and Evolutionary Stages. *J. theor. Biol*., 187, pp- 595-600.

Skinner, B.F. (1974). *About Behaviorism.* Vintage Books Edition, 1976, 291 pps.

Smith, J.D., Crossley, M.J., Beran, M.J. and Boomer, J. (2010). Implicit and Explicit Category Learning by Macaques (*Macaca mulatta*) and Humans (*Homo sapiens*). *Journal of Experimental Psychology: Animal Behavior Processes,* Vol. 36, No.1, pp. 54-65.

Sneddon, L.U. (2004). *Brain Research Reviews*, 46, pp. 123-130.

Sneddon, L.U. (2011). Pain Perception in Fish. Evidence and Implications for the Use of Fish. *Journal of Consciousness Studies*, 18, No. 9-10, pp. 209-229.

Sneddon, L.U., Braithwaite, V.A. and Gentle, M.J. (2003). Novel object test: Examining nociception and fear in in the rainbow trout. *J. Pain*, 4, pp. 431-440.

Sporns, O., Almássy, N. and Edelman, G.M. (2000). Plasticity in Value Systems and Its Role in Adaptive Behavior. *Adaptive Behavior*, 8, 2, pp. 129-148.

Stafford, T. (2014). The perspectival shift: how experiments on unconscious processing don't justify the claims made for them. *Frontiers in Psychology*, Vol. 5, Article 1067, pp. 1-4.

Starzak, T. (2016). Interpretations without justification: a general argument against Morgan's Canon. *Synthese,*

Stoerig, P. & Cowey, A. (2007). Blindsight. *Current Biology*, Vol. 17, No. 19, pp. 1-4.

Strawson, G. (1994). *Mental Reality.* MIT Press, Bradford Books.

Strawson, G. (2006). Realistic monism – why physicalism entails panpsychism. *Journal of Consciousness Studies*, 13, 10-11, pp. 3-31.

Suddendorf, T., Addis, D.R. and Corballis, M.C. (2009). Mental time travel and the shaping of the human mind. *Phil. Trans. R. Soc. B.*, 364, pps. 1317-1324.

Sutherland, M.A. & Tucker, C.B. (2011). The long and short of it: A review of tail docking in farm animals. *Applied Animal Behaviour Science*, 135, pp. 171-191.

Tartaglia, J. (2016). *Philosophy in a Meaningless Life*. Bloomsbury Publishing, 218 pps.

Thorndike, E.L. (1898). Animal Intelligence: An Experiential Study of the Associative Processes in Animals. *Psychological Monographs*, 2, 109.

Tinbergen, N. (1963). On aims and methods of ethology*. Zeischrift für Tierpsychologie*, 20, pp. 410-433.

Tononi, G. & Koch, C. (2015). Consciousness: here, there and everywhere? *Phil. Trans. R. Soc. B*, 370: 20140167, pp. 1-18.

Vandekerckhove, M. & Panksepp, J. (2011). A neurocognitive theory of higher mental emergence: From anoetic affective experiences to noetic knowledge and autonoetic awareness. *Neuroscience and Biobehavioral Reviews*, 35, pp. 2017-2025.

Vandekerckhove, M., Bulnes, L.C. and Panksepp, J. (2014). The emergence of primary anoetic consciousness in episodic memory. *Frontiers in Behavioral Neuroscience*, Vol. 7, Article 210, pp. 1-8.

Weary, D.M., Braithwaite, L.A. and Fraser, D. (1998). Vocal responses to pain in piglets. *Applied Animal Behavior Science*, 56, pp. 161-172.

Wegner, D.M. (2002). *The Illusion of Conscious Will*. Cambridge, Ma, MIT Press.

Whalen, P., Rauch, S., Etcoff, N., McInerney, S., Lee, M. and Jenike, M. (1998). Masked Presentations of Emotional Facial Expressions Modulate Amygdala Activity without Explicit Knowledge. *The Journal of Neuroscience, 18 (1),* 411–418.

White, N.M. (1989). Reward or reinforcement: What's the difference? *Neuroscience & Biobehavioral Reviews*. Vo. 13, 2-3, pp. 181-186.

Zehr, E.P. & Stein, R.B. (1999). What functions do reflexes serve during human locomotion? *Progress in Neurobiology*, Vol. 58, pp. 185-205.

Zola-Morgan, S. & Squire, L.R. (1984). Preserved learning in monkeys with medial temporal lesions: Sparing of motor and cognitive skills. *Journal of Neuroscience*, 4, pp. 1072-1085.

Åsli, O. & Flaten, M.A. (2012). In the Blink of an Eye: Investigating the Role of Awareness in Fear Responding by Measuring the Latency of Startle Potentiation. *Brain Sci*., 2, pp. 61-84.