Could Robots Feel Pain?

How Can We Know?

Bruce MacLennan
University of Tennessee, Knoxville, USA

ABSTRACT

This chapter considers the question of whether a robot could feel pain or experience other emotions and proposes empirical methods for answering this question. After a review of the biological functions of emotion and pain, the author argues that autonomous robots have similar functions that need to be fulfilled, which require systems analogous to emotion and pain. Protophenomenal analysis, which involves parallel reductions in the phenomenological and neurological domains, is explained and applied to the "hard problem" of robot emotion and pain. The author outlines empirical approaches to answering the fundamental questions on which depends the possibility of robot consciousness in general. The author then explains the importance of sensors distributed throughout a robot's body for the emergence of coherent emotional phenomena in its awareness. Overall, the chapter elucidates the issue of robot pain and emotion and outlines an approach to resolving it empirically.

Keywords: Consciousness, David Chalmers, Emotion, Feeling, Hard Problem, Neurophenomenology, Pain, Phenomenology, Protophenomena, Qualia

INTRODUCTION

As a society, we are already integrating machines into our bodies, and our ability to do this successfully grows daily, with true cyborgs as a seemingly inevitable end point. Moreover, we are progressing slowly but steadily toward autonomous robots whose behavior will give evidence of sentience. While it may be some time before such robots have intelligence comparable to ours, their eventual existence will raise ethical issues. Moreover, we will face these issues even before robots reach the level of human intelligence. As we have ethical standards for the treatment of laboratory animals, such as rats, so we will face ethical dilemmas in the treatment of autonomous robots with comparable mental capacities. Our notions of cruelty and ethical treatment of other beings depend to a large degree on their capacity to feel pain and to suffer in other ways: to feel fear, distress, anxiety, anguish, sorrow, loneliness, and loss. But we suppose there is a fundamental difference between *actually* experiencing these things (as we do) and *acting as though* we are experiencing them (as most people suppose machines to do). Even in the case of cyborgs, we would like a scientific basis for predicting the effects on an animal's sentience resulting from integration with artificial devices. These developments demand that we give our attention to long standing issues in the relation of mind and matter, that we move them from philosophical quandaries to practical ethical and ultimately legal issues. It is time to take them seriously and to develop methods to answer them reliably. After we understand robot sentience better, we will be in a position to address the ethical issues, but they are beyond the scope of this chapter.

Therefore this chapter focuses on robot consciousness, but there are several ways the word "consciousness" can be used in philosophy (e.g., Block, 1995). One sense has been termed *functional* or *access consciousness*, which includes *self-consciousness* — having an internal representation of the self about which one can reason — and *monitoring consciousness* — internal scanning and higher-order representation of mental state (Gutenplan, 1994, pp. 213–216). Of similar character is the *passive frame theory* of Morsella, Godwin, Jantz, Krieger, and Gazzaley (2015), which explains consciousness as "a frame that constrains and directs skeletal muscle output," a sort of clearinghouse between proposed actions and (unconscious) action deciders. These are important ideas with relevance to autonomous robots. In this chapter, however, the focus is on *phenomenal consciousness*, that is, the subjective experience of being a sentient being, of being aware, of feeling as opposed to reacting. In particular, we address the issue of whether a robot could feel pain or experience other emotions.

The principal problem of (phenomenal) consciousness is "to understand the relation between our subjective awareness and the brain processes that cause it" (MacLennan, 1995). This is commonly known as the *hard problem of consciousness*:

> *The really hard problem of consciousness is the problem of experience. When we think and perceive, there is a whir of information-processing, but there is also a subjective aspect. … It is widely agreed that experience arises from a physical basis, but we have no good explanation of why and how it so arises. Why should physical processing give rise to a rich inner life at all? It seems objectively unreasonable that it should, and yet it does. (Chalmers, 1995)*

More succinctly, "The hard problem of consciousness is the problem of explaining why any physical state is conscious rather than nonconscious" (Weisberg, 2012). This chapter addresses the hard problem in the context of robot pain and emotion. It argues that a robot can be programmed to exhibit emotional behavior (and that this is a useful thing to do), but this raises the question of whether such a robot would *feel* its emotions (MacLennan, 2014). In particular, could it feel pain? Is this possible? Under what conditions? Our goal is to approach these questions scientifically, that is, by means of hypotheses that can, in principle, be either confirmed or refuted empirically.

Some people might wonder why we should care whether a robot can feel emotions, or whether the question is even meaningful. Obviously it would be meaningful to the conscious robot itself, but it is also relevant to our treatment of the robot, for our notions of ethical treatment of others depend on their capacity to feel and especially to suffer. We send our old cars to the metal crusher without any feelings of sympathy or guilt, but whether we will do the same with our old robots will depend on whether we think they are capable of suffering, on whether we believe there is "anyone home."

Second, whether future autonomous robots are capable of feeling their emotions may affect their treatment of us. If they are incapable of feeling their emotions, then they will be incapable of truly empathizing with us; their understanding of us will be purely intellectual. Without too much exaggeration, their psychology could be characterized as that of a sociopath: calculating but unfeeling.

Finally, even if we never create robots capable of genuine feeling, it is an important question to address, for it sharpens our understanding of sentience in humans and other animals. Unless we can give a scientific answer to the question of possible robot consciousness, we cannot claim to understand our own minds very well. As Herb Simon famously said:

> *Perhaps the greatest significance of the computer lies in its impact on Man's view of himself … [T]he computer aids him to obey, for the first time, the ancient injunction "Know thyself." (Simon, 1977)*

BACKGROUND

Biological Functions of Emotion and Pain

Although it is becoming more widely accepted that some robots should respond to emotions and exhibit emotional behavior, some people question why we would ever want emotional robots. To understand their utility, it will be worthwhile to look at the functions of biological emotions (for more, see MacLennan, 2009, 2014).

What is an emotion? Rolls (2007) defines it as a state that is elicited by either the delivery or omission of a reward or punisher, which may be either present or remembered, and that functions as positive or negative reinforcement (thus inclining toward future approach or avoidance). Thus emotions function both as primary motivating mechanisms and as mechanisms for learned behavior and adaptation, both of which are important for autonomous robots.

Emotions are widespread in the animal kingdom, especially in the more complex species, such as mammals. Therefore, the adaptive advantages of emotions must far outweigh their disadvantages. What functions do they fulfill? Rolls (2005, 2007) has listed some of the most important, most of which are also relevant for future autonomous robots: Emotions elicit autonomic and endocrine responses that ready the animal for appropriate action. Emotions facilitate behavioral flexibility by implementing a “bow tie” information flow: various stimuli can elicit the same emotional response (e.g., fear), and this emotional response can lead to a variety of relevant responses (Rolls, 2006). Emotions are inherently motivating and maintain a persistent effect on cognitive processing, providing a coherent context so long as the eliciting situation continues. Moreover, because emotions are inherently important, they facilitate memory storage and retrieval, thus leading to more adaptive behavior. Finally, emotions facilitate intraspecies and interspecies communication and, in social species, they facilitate bonding. Of course, emotions do not always succeed in improving adaptation, but their prevalence in the animal kingdom shows that they are valuable behavioral adaptations, and this suggests that they will be similarly valuable in autonomous robots.

Many emotions are elicited by external sensory perception, but the *homeostatic* (or *primordial*) emotions are elicited by *interoceptors*, which sense the body’s state (Craig, 2003; Denton, 2006), and we expect future robots to have analogous internal sensors. The function of interoceptors is to maintain the integrity of the body and its physiological processes (i.e., homeostasis); they serve such emotions as hunger, thirst, “air hunger,” fatigue, and pain. Pain in particular results from excessive stimuli, which may indicate damage or potential damage to the body, and these stimuli are detected by specialized receptors called *nociceptors*. Some of the nociceptors are on the surface of the body and detect excessive temperature and pressure; others are embedded in muscles, tendons, and other internal structures, and respond to stretching, inflammation, and other conditions indicative of abnormal conditions. Information from these specialized interoceptors is transmitted on specialized nerve fibers (quickly on Aδ fibers, which mediate sharp pain, more slowly on C fibers, which mediate burning pain). These signals enter the spinal chord, and propagate to regions in the brain stem, thalamus, and somatosensory cortex.

The biological function of pain is apparently to motivate the animal to avoid potential damage to its body, to prevent either injury or interference with healing. Pain also enhances learning so that the animal is less likely to risk similar damage in the future. In short, pain has survival value, and people without the ability to feel pain (*congenital analgesia*) often injure themselves, for example, by burning themselves and by overstressing joints. Likewise, without a pain system — which serves to prevent damage or avoid further damage — robots are at risk of damaging themselves, as contemporary robots sometimes do.

Robot Pain and Other Emotions

Many of the biological functions of pain and other emotions, which Rolls enumerated, transfer directly to future autonomous robots. This is apparent if we imagine, for example, a cohort of autonomous robots cooperatively exploring a planet or engaging in a military operation. Although they might not have endocrine systems, per se, they could well have systems fulfilling analogous functions. For example, in an emergency they might shut down inessential subsystems, increase clock rates on more important ones, engage emergency power sources, deploy protective devices, hibernate, etc. Moreover, synthetic emotions can be expected to enhance cooperation among social robots, and especially to facilitate communication and cooperation with humans (Breazeal, 2003; Breazeal, Brooks, Gray, Hoffman, Kidd, Lee, Lieberman, Lockerd & Chilongo, 2004).

A pain response has obvious applications in robotics. Sensors monitoring temperature on its surface and in its interior, battery levels, torques on manipulators, excessive light on optical sensors, indicators of faults, dropped data, memory failures, etc. all generate signals analogous to pain, in that they should trigger high-priority interruption of other activities in order to eliminate the pain-inducing condition. Information coming in on the pain channels is implicitly high priority demanding real time response.

Such robots might act *as though* they are in pain, but would they actually *feel* pain? They might act afraid, but would they feel fear? Most people are inclined to answer "no" to both questions, and perhaps they are correct, but how can we know? In other words, is it possible for robots to have the information structures and control processes to implement pain and other emotions, yet have no internal experience of them, no "feelings"? In the context of the hard problem of human consciousness, philosophers call this a "zombie problem" (Campbell, 1970; Kirk, 1974; Kripke, 1980). Given our present understanding of human consciousness, there seems to be nothing contradictory in the idea of zombie humans, whose bodies have all the same physiological processes as normal humans, but who have no internal experience of subjective awareness (phenomenal consciousness). We have no complete and coherent theoretical account of how to get from the one to the other. The phenomenological properties of subjective awareness cannot be defined in terms of the physical predicates that describe these neurophysiological processes (Strawson, 1994, 2006). This *explanatory gap* is the root of the hard problem. Until we understand why such zombies are impossible, or under what conditions they are impossible, our understanding of the relation of physical and mental processes is radically incomplete, and there is a substantial hole in our scientific worldview. The situation with robot consciousness is similar, but people tend to be more skeptical of its possibility due to the differences in the underlying physical processes and the lack of direct experience or reliable reports (from robots). However, proposed answers to the question of robot consciousness are matters of opinion if offered in the absence of principled, preferably scientific, justification. How can we address these questions scientifically, that is, empirically?

PROTOPHENOMENAL ANALYSIS OF ROBOT PAIN AND EMOTIONS

The problem of robot consciousness is complicated by three factors. First, there are presently no robots so sophisticated that we might be tempted to call them conscious. Second, we have no experience of what it might be like to be a robot. Third, current and likely future robots are based on very different physical principles from animals, and therefore it is difficult to extrapolate from biological consciousness to robot consciousness. Therefore we need to investigate the relation between the structure of conscious experience and the structure of the physical processes from which it arises. By understanding the essential relations between consciousness and physical processes, we will be able to make informed judgments about whether very different physical systems, such as those in robots, might support conscious experience, including pain. We turn to neurophenomenological analysis as a research framework for addressing this relationship.

Overview of Neurophenomenological Analysis

We use *phenomenon* in a broad sense to refer to anything that appears (Grk., *phainomai*) in consciousness, including sensations, perceptions, hallucinations, dreams, recollections, intentions, mental discourse, emotions, and imagination; generally, for anything with a definite subjective "shape," but not including more diffuse and indeterminate moods, dispositions, etc. *Phenomenology*, then, is the study of the structure (*logos*) of possible phenomena, that is, the science of consciousness. There are, of course, many approaches to phenomenology, but I use the term broadly to refer to the Husserl – Heidegger – Merleau-Ponty philosophical spectrum, which uses first-person methods to investigate the structure of human experience, understood as sentient, embodied, situated, and purposeful. In particular, *experimental phenomenology* uses systematic, introspective techniques for the empirical investigation of the structure of conscious experience from the *inside*, that is, from the perspective of the conscious subject (Ihde, 1986; McCall, 1983). Empirical science must embrace phenomenological methods in order to achieve a comprehensive understanding of nature and, in particular, of human nature (MacLennan, 2015).

The hard problem, then, can be addressed by a process of *neurophenomenological analysis*, which proceeds by parallel reductions in the neurological and phenomenological domains, that is, by analyzing consciousness from both exterior and interior perspectives (Laughlin, McManus & d'Aquili, 1990; Lutz & Thompson, 2003; Rudrauf, Lutz, Cosmelli, Lachaux & Le Van Quyen, 2003; Varela, 1996).

Neurological reduction is familiar: we seek to understand large-scale neurophysiological processes in terms of smaller ones, reducing, for example, processes in the cortex to processes in individual neurons, and these to smaller scale biological, and ultimately physical, processes. This can be termed an *external* reduction, since we are investigating the nervous system from the outside, as an object, in the normal scientific way. In parallel, however, we can conduct an *internal* reduction, in which we use phenomenological techniques to investigate consciousness from the inside. This is accomplished within the phenomenological domain by both qualitative and quantitative reductions.

In a *qualitative reduction* the experimental phenomenologist analyzes phenomena into smaller phenomena of qualitatively different kinds. This is simplest to understand in the realm of sensory experience, for visual experience is different from auditory experience, which is different from olfactory experience, and so forth. In the parallel neurological reduction, these phenomenological modalities correspond to various sensory areas in the brain. Accurate phenomenology recognizes, however, that these modalities are not completely independent, for visual perception can affect auditory perception, and vice versa. This is confirmed, in the parallel neurological domain, by neurons in auditory cortex that respond to visual stimuli and vice versa.

Other qualitative modalities can be identified through phenomenological analysis; they include imagination, inner discourse, recollection, dreaming, waking, anticipation, desire, fear, rage, etc. The challenge for phenomenology is to identify these various modalities and to classify them carefully, since their relations are complex and subtle. We guard against naïve introspection by conducting carefully controlled phenomenological experiments, supported by training in phenomenological methods, and by insights from neuroscience that suggest parallel phenomenological hypotheses to test. Conversely, phenomenological observations suggest parallel neuroscientific hypotheses.

In contrast to qualitative reduction, which analyzes phenomena into smaller phenomena of *different* kinds, *quantitative reduction* analyzes a phenomenon into smaller phenomena of the *same* kind. The simplest application of quantitative reduction is in visual experience, for the visual phenomena can be analyzed into smaller patches of visual experience. As a first approximation, we might think of elementary patches of color and intensity analogous to the pixels in a digital image. However, knowledge of neurophysiology reveals that this is an oversimplification, for retinal "spot detectors" are overlaid by center-surround filters, edge detectors, Gabor filters, and so forth, in a complex bidirectional processing hierarchy that affects even "raw" visual experience (e.g., Daugman, 1993; De Valois & De Valois, 1993). Therefore, even for visual phenomena, where a quantitative analysis might seem straight-forward, we find that there are subtle effects that might be overlooked in the absence of parallel neurological investigations.

Haptic and proprioceptive phenomena provide another example of comparatively simple quantitative reduction. Sensory neurons respond to relatively small patches of skin and are mapped systematically in sensory cortex. In this case we can find relatively simple correspondences between conscious haptic experiences and neural processes. The receptive fields of somatosensory neurons correspond closely to conscious sensory experience. Topographic maps in other sensory areas (for example, tonotopic maps in auditory cortex) suggest other quantitative phenomenological reductions (MacLennan, 2010).

As previously mentioned, sensory experience is only one small component of conscious experience, and so neurophenomenological analysis of consciousness cannot be limited to sensation. Rather, it must include parallel neurological and phenomenological reductions of all the mental phenomena mentioned above, and many more as well, that is, reductions for the totality of conscious experience. Therefore, completion of this research program depends on future progress in neuroscience, as well as on a comprehensive research program in experimental phenomenology. The ultimate goal of neurophenomenological research is a comprehensive analysis of the structure and dynamics of conscious experience in terms of more elementary phenomenological processes that can be correlated with neural processes. Such an understanding of the physical basis of human consciousness will provide a basis for determining the physical preconditions for robot consciousness.

Protophenomena and Activity Sites

Since future autonomous robots will probably have a very different physical basis from animals, we need to explore the neurophenomenology of human consciousness in more detail so that we can see the essential issues in robot consciousness. Although this neurophenomenological research program is very far from completion, we can get some insights into the hard problem by anticipating its final form. Reduction can proceed only so far. On the neurological side, there are smallest units with direct relevance to consciousness, but it is premature to say what they are. Defensible candidates include minicolumns, neurons, synapses, and neurotransmitter receptor sites. The parallel reduction on the phenomenological side proceeds as far as corresponding elements of the conscious state, which may be termed *protophenomena* (Chalmers, 1996, pp. 126–7, 298–9; Cook, 2000, 2002a, 2002b, chs. 6–7, 2008; MacLennan, 1996a; cf. *proto-qualia* in Llinas, 1988; *phenomenisca* in MacLennan, 1995). Williams James referred to them as "mental atoms" and "aboriginal atoms of consciousness" (James, 1890/1955, vol. I, ch. 6, p. 149). We may define protophenomena as the smallest constituents of conscious phenomena. Phenomena, then, as the determinate contents of consciousness, consist in coherent assemblies of their constituent protophenomena. Since protophenomena have (as yet imperfectly defined) elementary subjectivity, phenomenological properties (properties of conscious experience) will be explainable in terms of the properties of protophenomena. At the bottom of the parallel phenomenological and neurological reductions we have protophenomena and their corresponding simple neural processes. At this time, I believe that their correlated physical and phenomenal properties must be accepted as a brute fact of nature, but protophenomenal properties are likely to be simple (as explained below), and so there is hope we may eventually explain this correlation.

Progress in neurophenomenology has not progressed so far that we can identify the neural processes associated with protophenomena, and so, in the absence of a more specific identification, we refer to them as *activity sites*. We use this intentionally vague term because the activity sites might be minicolumns, neurons, synapses, neurotransmitter receptors, or something else. In any case, it must be stressed that protophenomena are very small compared with phenomena. If activity sites are minicolumns, for example, then the conscious state might comprise 10^8 protophenomena; if the activity sites are synapses, then the number is more like 10^{15}. In terms of scale, the relation between protophenomena and phenomena is analogous to the relation between atoms and macroscopic objects.

The vast difference in scale between phenomena and protophenomena in humans has important consequences, for the typical contribution of each protophenomenon to the conscious state is miniscule. Therefore, although protophenomena are the constituents of phenomena, it is misleading to think of them as very small phenomena (they are *proto*phenomena, not phenomena). While protophenomena have the property of elementary subjectivity, which allows them to combine to constitute a subjective state, we are not conscious of protophenomena per se. This seems paradoxical, so an analogy may help. An individual H_2O molecule is not liquid; nevertheless, these molecules have the physical properties from which liquidity emerges when enough H_2O molecules are combined (at appropriate temperatures and pressures). The change of one protophenomenon will not usually affect a phenomenon, qua phenomenon, just as the addition or removal of an H_2O molecule will not change a water drop, qua water drop. To put it in behavioral terms, a person is not usually able to report a change in a single protophenomenon. This is one of the facts that complicates empirical investigation of the relation of protophenomena and activity sites.

Protophenomenal Intensity and Interdependencies

We have defined protophenomena as the elementary constituents of the conscious state, but have not said anything about their properties and how those properties might be related to physical processes at the corresponding activity sites. If we think of protophenomena that are conceptually simple, such as patches of color intensity in the visual field, or sensations of pressure on patches of skin, then we see that they have a degree of presence in the conscious state. In these cases, there is corresponding activity in neurons (in the retina, LGN, sensory cortices, etc.). Therefore, at least to a first approximation, each protophenomenon has an *intensity*, which represents its degree of presence in the conscious state. As a consequence, the totality of protophenomena correspond to the elementary degrees of freedom of the conscious state. That is, at any given time, the joint intensities of a person's protophenomena constitute that person's conscious state at that time. The protophenomena themselves (i.e., independent of their intensities) and the interdependencies among them (explained below) define the set of a person's possible conscious states and the possible sequences of states. Thus, the protophenomena and their interdependencies define the structure of a person's phenomenological world. We would expect the same to be true of a robot if its basic physical processes support protophenomena; that is, the interdependencies among the robot's protophenomena would determine the structure of its conscious states.

Protophenomenal intensity is correlated with some physical quantity at the corresponding activity site, which we call its *activity*, but the nature of this activity will depend on what the activity sites turn out to be. For example, if the activity sites are neurons, then the activity correlated with protophenomenal intensity might be firing rate, membrane potential, number of occupied receptor sites, or something else associated with the neuron as a whole. Determining the physical processes associated with protophenomenal intensity will require complex neurophenomenological experiments, but techniques such as optogenetics are bringing them within the range of feasibility. For example, by making physical interventions in potential activity sites for especially salient protophenomena (by opening or closing specific ions channels, for example) and having the subjects observe and report on changes in their conscious states, we can narrow down the nexus between protophenomenal intensity and physical processes at the corresponding activity sites. This will help us to determine the sorts of physical processes that support protophenomena, and therefore potential physical substrates for robot consciousness.

Since the only property of protophenomena is their intensity, we need to explain what distinguishes visual protophenomena, from acoustic protophenomena, from pain protophenomena, and so forth. Neuroscience suggests an answer, for sensory and motor areas are organized in topographic maps, which systematically represent the topology of a sensory or motor domain. These abstract topologies are represented in the spatial organizations of the maps and the resulting patterns of interconnections among their neurons. Therefore, the neurons represent related points in these topologies by virtue of their interconnections (there is no difference between a neuron in auditory cortex representing Middle-C and a neuron in visual cortex representing a patch of red color). That is, the populations of neurons in these various sensory and motor areas are essentially identical (as are their local circuits within minicolumns); their "meaning" arises from connections between minicolumns in the maps and from connections between maps. These observations suggest (via the parallel reductions) that essentially identical protophenomena acquire their qualitative aspect by virtue of interdependencies with other protophenomena (which correspond to physical connections among activity sites). This provides a basis for giving a qualitative "shape" to phenomena in robot consciousness.

In general terms, we may say that the subjective dynamics of the protophenomena is correlated with the neurodynamics of the activity sites, which jointly constitute the neurodynamics of the brain, which is in turn governed by the physical interconnections among the activity sites. Global neurodynamics emerges from the coordinated activity of interconnected activity sites. Correspondingly, the coherent dynamics of protophenomenal intensities constitutes the emergence of macroscopic phenomena in consciousness. Corresponding to the connections between activity sites are the interdependencies among protophenomena, by which the evolving intensities of protophenomena mutually govern the intensities of other protophenomena. Depending on our assumptions about the identity of the activity sites, one can formulate differential equations for protophenomenal intensities (MacLennan, 1996b).

Protophenomena are not qualia, per se; rather, the protophenomenal hypothesis explains qualia in terms of protophenomenal interdependencies. This is motivated by neuroscience, for while there are many types of neurons, we find a common architecture across much of the cerebral cortex. The same kinds of neurons serve visual perception as serve auditory perception. Indeed, this has been demonstrated by causing nerve fibers carrying visual information to grow into auditory cortex, thus "reprogramming" auditory cortex so that it supports visual perception (Sur, 2004). These experiments imply that, on the neuroscience side of the parallel reductions, perceptual properties are determined by neural connections (more generally, by physical dependencies between activity sites). Correspondingly protophenomena acquire their qualities — and therefore constitute qualia — by virtue of their interdependencies. This does not preclude the possibility that there are several kinds of protophenomena, corresponding to kinds of neurons, or that there are different kinds of protophenomenal dependencies, corresponding to different neurotransmitters (e.g., especially important for pain transmission are the neurotransmitter *glutamate* in the Aδ fibers and the neuropeptide *substance P* in the C fibers), but the simpler hypothesis is that these distinctions are inessential. In an analogous manner, we expect robot qualia — fundamental subjective qualities in a robot's consciousness — to be a function of its protophenomenal interdependencies, which might be quite different from those of humans and other animals, and therefore create a different kind of conscious experience.

The Neurophenomenology of Pain and Other Emotions

With this background, we proceed to apply neurophenomenological analysis to emotion, and in particular to pain. Since animals are the only things we know of that feel pain and other emotions, we must begin with them as a basis for exploring the possibility of emotions and pain in future robots; we have to understand the physical basis of pain in animals before we can understand it in robots. The first task is a phenomenology of emotion, already a complex undertaking (MacLennan, 2009). Even the definition of "emotion" is contentious. For example, Plutchik (2003, pp. 18–19) lists twenty competing definitions, and Kleinginna and Kleinginna (1981) present a taxonomy of ninety definitions. The structure of emotional experience is even more complex, but a consensus is emerging from multiple studies (Prinz, 2004, p. 90). One proposed structure is a *circumplex* of emotions, which maps emotions onto a cone (Plutchik, 2000; Plutchik & Conte, 1997). Eight positions around the circumference define *primary emotions*, which come in four opposed pairs reflecting *valence* (whether the emotion is positive or negative). *Secondary emotions* are mixtures of the primary emotions (like mixtures of primary colors), and emotional *intensity* is measured by distance from the cone's apex (where all emotions have zero intensity), analogous to brightness in color. The phenomenology of pain is simpler, but still more complex than commonly supposed (discussed below).

It is necessary to distinguish *pain*, as raw sensation, from *suffering*, which is an affective-cognitive phenomenon. Similarly, *emotion* must be distinguished from *feeling*, for an organism may respond emotionally without the consequent conscious experience of that reaction as an emotion (Damasio, 1999, pp. 42–49). To explain the neurophysiology of emotion, we can begin with the *somatic feeling theory* of emotion, which is also known as the *James-Lange theory*, for it was first proposed independently by William James in 1884 and by Carl Lange in 1885 (James, 1884; Lange, 1885). Although this theory was proposed more than a century ago, most contemporary theories are some modification of it (e.g., Damasio, 1994, 1999; Prinz, 2004). The essence of the theory is that the initial emotional response is unconscious, and that subsequent conscious experience of the emotion is a result of sensing changes in the body resulting from the unconscious reaction. For example, a threatening perception may be processed unconsciously and lead to somatic effects, such as tensed muscles and increased heart and breathing rates (Plutchik, 2003, p. 127). Then, as a consequence of conscious processing of the stimulus, these somatic changes are experienced as fear. That is, we consciously experience our somatic reaction and perceive the frightening stimulus. Similarly, when dealing with robot emotions, we will need to distinguish the conscious and somatic aspects.

Contemporary theories of the neurophysiology of emotion recognize three levels of processing (Prinz, 2006, ch. 9). Damasio (1999, p. 8) has characterized them as "an emotion, the feeling of that emotion, and knowing that we have a feeling of that emotion." Similarly, pain has three dimensions: sensory-discriminative, affective-motivational, and cognitive-evaluative (Melzack & Casey, 1968). The *sensory-discriminative* aspect of pain has dimensions of bodily location, intensity, duration, and quality (heat, cold, pressure, etc.); the *affective-motivational* aspect relates to the experience of unpleasantness and the urge to eliminate the source of pain; the *cognitive-evaluative* aspect relates to conscious appraisal of the pain and the response to it (e.g., "I burned myself, and need to get the burn ointment"). The three dimensions are mutually determinative, so there are top-down as well-as bottom-up influences (mediated by descending and ascending neural pathways). Therefore, the cognitive-evaluative processing can modulate both affective-motivational and sensory-discriminative processing. For example, our appraisal of the burn and its severity can influence its perceived unpleasantness and its intensity.

At the lowest level of the emotional hierarchies are somatic neurons responding to somatic conditions. Prinz (2006, ch. 9) says these are neurons located in primary somatosensory cortex, pons, and insula with small receptive fields responding to visceral organs, skeletal muscles, hormone levels, etc. In the case of pain, nociceptors, which respond to excessive pressure, stretching, temperature, irritants, etc., send Aδ and C fiber projections to the spinal chord, where they synapse on second-order neurons, which ascend into the brainstem (parabrachial nucleus) and insula. These low-level neurons correspond to the activity sites of emotional protophenomena, which are the elementary constituents of pains and other emotional phenomena (feelings), but are not conscious pains or feelings per se. Robots might have similar sensors for important conditions in their bodies.

The next level of emotional processing takes place at the level of cortical maps, especially in secondary somatosensory, dorsal anterior cingulate, and insular cortices (Prinz, 2006, ch. 9), which organize these neural responses in terms of bodily location and kind of stimulus on the basis of short-range connections within these maps and longer-range connections to other maps. Pain is also mapped into bodily location at this level, precisely in the case of superficial pain, more vaguely in the case of visceral pain. Pain signals are somatotopically mapped in the posterior ventral medial nucleus of the thalamus (Craig, 2003) Interdependencies among the corresponding protophenomena (bodily mapped and multimodal) cause them to cohere into emotional phenomena, that is, into consciously experienced emotions. The dependence on topographic location is apparent from the phenomenon of *referred pain*, in which cortical neurons that serviced an amputated limb reassign themselves to other body locations that are serviced by nearby neurons in the map (e.g., Karl, Birbaumer, Lutzenberger, Cohen, & Flor, 2001). If robots similarly map the bodily location of their interoceptor responses and integrate them into feelings, then they might experience pain and other emotions from this level of processing.

The third level of processing, which takes place in higher cortical regions, possibly in ventromedial prefrontal cortex and rostral anterior cingulate cortex (Prinz, 2006, p. 214), is where, as an integrated sentient being, we consciously recognize that we are having a feeling and give it a name: "My foot hurts!" "Ouch, that's hot!" "I'm afraid!" Corresponding to this level are cognitive phenomena arising from protophenomena whose interdependencies give them this cognitive character. Robots might have these feelings as well, if they have these higher levels of cortical processing.

The foregoing is just an outline. Understanding conscious experience of emotions will require a detailed neurophenomenology, which depends in turn on the ongoing neurophysiological investigation of emotion, but especially on more systematic phenomenological investigation of the human experience of emotion. The case of pain is simpler, but the difficulty people have in describing pains precisely, and sometimes even in localizing them, illustrates that the phenomenology of pain is far from trivial. As in other empirical sciences, training is required for accurate experimental phenomenology.

Nonbiological Protophenomena?

We can apply the foregoing summary of the neurophenomenology of human emotion to the question of robot feelings. This will highlight empirical questions that need to be answered before we can give definitive answers to the question of robot consciousness, and in particular, robot pain.

Due to the conceptual difficulties in reducing phenomenological properties to physical properties or vice versa, protophenomenal theory generally treats the correlation between protophenomenal intensity and the physical activity at activity sites as a brute fact of nature. Therefore, protophenomenal theory can be classified as a kind of *dual-aspect monism* (Atmanspacher, 2012), which supposes that there is one sort of "stuff" in the universe, but that it has two mutually irreducible aspects, an external, physical aspect, and an internal, phenomenological aspect. (More specifically, protophenomenal theory can be classified as *Type-F monism*; see Chalmers, 2002.) To put it differently, the properties of a physical system are experienced differently depending on whether or not the experiencer *is* that system.

Nevertheless, this leaves open the question of whether all physical systems have these two aspects or only certain ones. In protophenomenal terms, what sorts of physical systems can be activity sites with associated protophenomena? This is an empirical question, but difficult to investigate. In fact, it is even an open question whether all neurons have protophenomena. Blind-sight, for example, reveals that visual processing can be unconscious as well as conscious (Weiskrantz, Warrington, Sanders, & Marshall, 1974). Certainly, much of the neural activity in our brains is unconscious, but that does not imply that the neurons involved lack protophenomena. It might be that the protophenomenal intensities are too diffuse or incoherent to constitute salient phenomena.

To answer these questions, we need to identify protophenomena that are individually sufficiently salient to have reportable effects in the conscious state, and then to intervene physically in the corresponding activity site to determine which physical variables affect protophenomenal intensity. In the case of pain, for example, we know that opioid analgesics act on neurons in the spinal chord and brain, either decreasing neurotransmitter release by blocking calcium ion intake, or inhibiting the postsynaptic neuron by opening potassium ion channels and raising its firing threshold. Optogenetics might be used to control these channels directly, thus teasing apart their effects on conscious pain perception. Techniques for this kind of very precise experimental intervention are steadily improving (e.g., Losonczy, Makara, & Magee, 2008; Petit, Wang, Gee, & Augustine, 1997; Service, 2013). Pain protophenomena are especially useful for neurophenomenological experimentation because they are relatively salient and isolable from other protophenomena.

Investigations of this kind will eventually reveal the physical structures and processes associated with protophenomena. Conceivably, they could be peculiar to biological systems (e.g., dependent on neurotransmitters and their receptors), in which case it would be unlikely that robots built on very different physical principles would have protophenomena, and therefore conscious states. On the other hand, it might turn out that many different physical processes support protophenomena, in which case robot consciousness becomes a real possibility, so long as the requisite physical processes occur in the robot. Note, however, that the mere presence of protophenomena is not sufficient for a robot to be conscious; their protophenomenal interdependencies would have to be structured in such a way that the protophenomena cohere into macroscopic phenomena (discussed below, **Protophenomenological Structure of Robot Emotion**).

These are open questions, but we can consider the implications of several hypotheses. First, Norman Cook (2000, 2002, 2008) has proposed an interesting hypothesis about activity sites: he argues that the intensity of a protophenomenon is correlated with the flux of ions across the neural cell membrane during the generation of an action potential. This can be considered an elementary act of perception in which the neuron senses the interneuronal environment. If this hypothesis is correct, what might it tell us about the essential physical properties of an activity site? Assuming it does not depend on biological specifics, it might be sufficient that (1) there is some sort of boundary separating the activity site from its environment, (2) that the activity site has the capability of sensing its environment, and (3) that its internal processing possibly results in some modification of its environment (i.e., a very simple sensory-motor loop). If these are, indeed, the essential characteristics of an activity site, then they are certainly implementable in future robots.

Chalmers (1996, ch. 8) presents a related hypothesis based on *information spaces* that have both physical and phenomenal aspects, the two together comprising an activity site with its protophenomenon, for information is simultaneously physical and phenomenal. The physical system has the structure sufficient to represent some simple information state and the causal relations to use this state to affect the physical states of some other physical systems. That is, it represents “differences that make a difference.” The associated protophenomenon has a corresponding formal structure and interdependencies connecting it to other protophenomena.

Combining the insights from Cook and Chalmers, we may consider the hypothesis that the neuron is the activity site. The information state is represented by the electrical state of the neuron (crudely, whether it is firing or not). The neuron senses its intercellular environment by means of its receptors, which respond to the chemical environment in the vicinity of the synapses. This increases the *system mutual information* between the cell state and its environment, in effect transferring information into the neuron (MacLennan, 2011). The entropy of the joint system decreases. Based on this input and the neuron's prior state, the neuron fires or not; that is, it makes a decision. Transmission of the action potential has physical effects on other neurons, muscles, glands, etc. The neuron, therefore, can be considered an information space with both physical and phenomenal aspects (i.e., it is an activity site together with its protophenomenon).

If Chalmer's hypothesis — that protophenomena and activity sites are the complementary aspects of information spaces — is correct, then we can conclude that robots could support activity sites and their associated protophenomena, for a robot's central processor certainly realizes information spaces. Again, we cannot conclude that such a robot is conscious, for protophenomena are necessary but not sufficient for consciousness. The protophenomenal interdependencies must to so structured that the protophenomena cohere into full-fledged phenomena. In particular, for emotional consciousness, the structure must be such so that they cohere into emotional phenomena (felt emotions). What is this structure?

Phenomenological Structure of Robot Emotions

To understand the structure of robot emotions, we can look at the functions of emotions in animals, which we reviewed above, and anticipate analogous functions in future robots. In brief, emotions are rapid evaluations of external or internal situations that require a prompt response. They ready the robot for action or information processing and may initiate those processes. Effects might include adjustment of clock rates, energy management, high-priority computation, deployment and priming of specialized sensors and actuators, etc. These are analogous to unconscious results of emotional processing in animals, but according to somatic feeling theory, felt emotions arise from higher order processing of somatic sensations. In robots, this would be based on interoceptors (internal sensors that monitor the robot's own state) and on direct signal connections carrying information about the effects of early emotional evaluation. In particular, some of these interoceptors could be analogous to pain receptors in that they monitor potentially damaging situations, such as excessive input to external sensors, excessive stresses on mechanical systems, dangerous electrical loads, excessive temperature, and physical damage. Some will trigger immediate "reflexive" reactions, such covering sensors, retracting actuators, shutting down low priority subsystems, or starting fans. Interoceptors in these systems might constitute activity sites with associated protophenomena (as previously discussed), but they might not rise to the level of conscious experience, which requires higher-level integration.

Many of these sensors and signal paths will be distributed around the robot's body and therefore will convey somatically structured information, with associated protophenomena. Higher order processing will integrate them into coherent somatic phenomena, that is, into felt pain and emotion. In sophisticated robots, further cognitive processing of this and other sensor information will allow the robot to draw conclusions about its emotional state, and if it is in pain, to decide how best to respond.

More specifically, if the interoceptors are activity sites, they will have a bodily organization, and therefore the protophenomena will constitute the elements of a phenomenological body (the experience of being an embodied consciousness). Each interoceptor has a response function defined over its input space, responding, for example, to high temperature or excessive torque. Therefore, its protophenomenon effectively represents this subspace, but it is unrelated to other protophenomena in the absence of interdependencies, which are mirrored in the connections between activity sites. The interconnection of these activity sites with each other and their integration into higher-order sensory areas stitches together a topology over their composite input space. This topology defines the structure of a space of corresponding emotional phenomena, in a phenomenological body (cf. Craig, 2003). This phenomenological structure will depend on the physical structure of the robot's body and on the organizational requirements for the robot's emotional subsystems.

One may wonder how similar a robot's emotions will be to our own. We can expect that some of a robot's somatosensory spaces will have a similar structure to ours. For example, like us they will have skin and joint sensors distributed throughout their bodies. Some of their interoceptors will be analogous to ours (sensor overstimulation, temperature, stress, strain, physical damage). In these cases, we can expect their emotional experiences will be similar to ours, for they will have a similar topology. We will be able to imagine their emotional experiences, as they may be able to imagine ours, just as we can imagine the haptic, muscular, and skeletal sensations of other mammals without too much trouble. On the other hand, we can anticipate that robot bodies will be radically different from ours in some ways, and these differences may be especially prominent in the emotion systems. For example, a robot is unlikely to have a heart or lungs per se, which are important to our somatic emotional responses, nor is it likely to have an endocrine system. It may have analogous interoceptor input (e.g., power and temperature monitors, early warning sensors), but the topology of their input spaces might be quite different from our interoceptors. In these cases, it might be more difficult for us to imagine a robot's emotional responses (just as it is difficult for us to imagine what it is like to be a reptile or crustacean). That is, the phenomenology of their emotions could be quite different to ours and unique to their "form of life." Although we may be able to understand their phenomenology intellectually, we might not be able to imagine it.

If robots can be conscious, but their emotional phenomena are very different from ours, should we call these phenomena "emotions" at all? Might this be an an overextension of the concept based on a loose analogy? The answer, I think, depends on the role played by these information processing systems within the robot. If they fulfill similar functions to emotions (as enumerated above) and fulfill them in similar ways, then I think it is useful to classify them as emotions. In particular, if they are states encoding goals of critical importance that are directly motivating and that have appropriate, persistent, and pervasive effects on the behavioral state of the robot, then they have much in common with emotions. The similarity is reinforced if they have effects on the robot's conscious state similar to our emotions' effects on our conscious state, for example, if they modulate the content and dynamics of the conscious state in a way that is relevant to the behavioral purposes that they serve. Certainly, a robot's conscious perception of impending damage to its structure could be reasonably termed "pain" if it leads to avoidance and negative reinforcement.

FUTURE RESEARCH DIRECTIONS

As the preceding sections reveal, the question of the possibility of robot pain and feeling cannot be answered at this time, but we have outlined research programs directed to this end. Whether we are considering future cyborgs, with artificial devices integrated in natural nervous systems, or completely artificial autonomous robots that exhibit sentience, we need a more principled understanding of the relation of conscious awareness and physical processes. By neurophenomenological analysis, we can reduce these issues to a simpler problem: understanding the relation between activity sites and their associated protophenomena. This problem is empirical and the requisite experimental techniques are being developed, but much research needs to be done before we can begin to determine necessary and sufficient conditions for physical systems to have associated protophenomena. Moreover, as we have explained, conscious experience requires more than just protophenomena; their independencies must be so structured that the protophenomena cohere into determinate phenomena. Understanding this process requires extensive neurophenomenology exploring the neurological correlates of conscious experience and its structure. Progress on the neurological side is progressing rapidly, but comparable progress on the phenomenological side will require more experimenters (and subjects!) trained in phenomenological techniques. Davidson's well-known fMRI investigations of expert meditators is an example of what can be done (e.g., Brefczynski-Lewis, Lutz, Schaefer, Levinson, & Davidson, 2007).

CONCLUSION

Under appropriate conditions, future robots could have conscious emotional experiences — including in particular, feeling pain — analogous, but not necessarily identical to these experiences in humans or other animals. Determining the precise conditions for robot emotional consciousness will require detailed and comprehensive neurophenomenological experiments. The central issues are to determine the range of physical processes with corresponding protophenomena (for these are the elementary constituents of consciousness) and to chart the neurophenomenological structure of emotion in humans and other animals. This is a complex, difficult, and long-range research program, but when it is completed it will provide a principled, evidence-based answer to the question of whether and under what conditions a robot could feel pain, suffering, and other emotions — perhaps even joy.

REFERENCES

Atmanspacher, H. (2012). Dual-aspect monism à la Pauli and Jung. *Journal of Consciousness Studies*, 19, 96–120.

Block, N. (1995). On a confusion about a function of consciousness, *Behavioral and Brain Sciences*, 18, 265–266.

Breazeal, C. (2003). Toward sociable robots. *Robotics and Autonomous Systems*, 42, 167–175.

Brefczynski-Lewis, J.A., Lutz, A., Schaefer, H. S., Levinson, D. B., & Davidson, R. J. (2007). Neural correlates of attentional expertise in long-term meditation practitioners. *Proceedings of the National Academy of Sciences*, 104, 11483–11488.

Breazeal, C., Brooks, A., Gray, J., Hoffman, G., Kidd, C., Lee, H., Lieberman, J., Lockerd, A., & Chilongo, D. (2004). Tutelage and collaboration for humanoid robots. *International Journal of Humanoid Robots*, 1(2), 315–348.

Campbell, K. K. (1970). *Body and mind*. New York, NY: Doubleday.

Chalmers, D. J. (1996). *The conscious mind*. New York, NY: Oxford University Press.

Chalmers, D.J. (2002). Consciousness and its place in nature. In D.J. Chalmers (Ed.), *Philosophy of mind: Classical and contemporary readings* (pp. 247–272). Oxford, UK: Oxford University Press.

Cook, N. D. (2000). On defining awareness and consciousness: The importance of the neuronal membrane. In *Proceedings of the Tokyo-99 Conference on Consciousness*. Singapore: World Scientific.

Cook, N. D. (2002a). Bihemispheric language: How the two hemispheres collaborate in the processing of language. In T. Crow (Ed.), *The speciation of modern Homo sapiens* (ch. 9). London, UK: Proceedings of the British Academy, v. 106.

Cook, N. D. (2002b). *Tone of voice and mind: The connections between intonation, emotion, cognition and consciousness*. Amsterdam, Netherlands: John Benjamins. doi:10.1075/aicr.47

Cook, N. D. (2008). The neuron-level phenomena underlying cognition and consciousness: Synaptic activity and the action potential. *Neuroscience*, 153(3), 556–570. doi:10.1016/j.neuroscience.2008.02.042 PMID:18406536.

Craig, A. D. (2003). Interoception: The sense of the physiological condition of the body. *Current Opinion in Neurobiology*, 13(4), 500–505. doi:10.1016/S0959-4388(03)00090-4. PMID 12965300.

Damasio, A. R. (1999). *The feeling of what happens: Body and emotion in the making of consciousness*.

New York: Harcourt, Brace & Co.

Daugman, J. G. (1993). An information-theoretic view of analog representation in striate cortex. In E. L. Schwartz (Ed.), *Computational Neuroscience* (pp. 403–423). Cambridge: MIT Press.

Denton, D. A. (2006). *The primordial emotions: The dawning of consciousness*. Oxford, UK: Oxford University Press.

De Valois, R. L., & De Valois, K. K. (1993). A multi-stage color model. *Color Vision*, 33, 1053–1065.

Gutenplan, S. (Ed.) (1994). *A companion to the philosophy of the mind*. Oxford: Blackwell.

Ihde, D. (1986). *Experimental phenomenology: An introduction*. Albany, NY: State University of New York Press.

James, W. (1884). What is an emotion? *Mind*, 9, 188–205.

James, W. (1955). *The principles of psychology, authorized edition*. New York, NY: Dover (Reprint of Henry Holt, 1890 edition).

Karl, A., Birbaumer, N., Lutzenberger, W., Cohen, L.G., Flor, H. (2001). Reorganization of motor and somatosensory cortex in upper extremity amputees with phantom limb pain. *The Journal of Neuroscience*, 21, 3609–18.

Kirk, R. (1974). Zombies versus materialists. *Aristotelian Society*, 48 (suppl.), 135–152.

Kleinginna, P. R., & Kleinginna, A. M. (1981). A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation & Emotion*, 5(4), 345–79.

Kripke, S. A. (1980). *Naming and necessity*. Cambridge, MA: Harvard University Press.

Lange, C. J. (1885). *Om sindsbevaegelser: Et psyko-fysiologisk studie*. Copenhagen: Jacob Lunds.

Laughlin, C. D., Jr., McManus, J., & d'Aquili, E. G. (1990). *Brain, symbol and experience: Toward a neurophenomenology of consciousness*. Boston, MA: New Science Library.

Llinas, R. R. (1988). The intrinsic electrophysiological properties of mammalian neurons. *Science*, 242, 1654–1664.

Losonczy, A., Makara, J. K., & Magee, J. C. (2008). Compartmentalized dendritic plasticity and input feature storage in neurons. *Nature*, 452(7186), 436–440. doi:10.1038/nature06725 PMID:18368112

Lutz, A., & Thompson, E. (2003). Neurophenomenology: Integrating subjective experience and brain dynamics in the neuroscience of consciousness. *Journal of Consciousness Studies*, 10(9/10), 31–52.

MacLennan, B. J. (1995). The investigation of consciousness through phenomenology and neuroscience. In J. King & K. H. Pribram (Eds.), *Scale in conscious experience: Is the brain too important to be left to specialists to study?* (pp. 25–43). Hillsdale, NJ: Lawrence Erlbaum.

MacLennan, B. J. (1996a). The elements of consciousness and their neurodynamical correlates. *Journal of Consciousness Studies*, 3(5/6), 409−424. Reprinted in J. Shear (Ed.), *Explaining consciousness: The hard problem* (pp. 249–266). Cambridge, MA: MIT Press, 1997.

MacLennan, B. J. (1996b). Protophenomena and their Neurodynamical Correlates. University of Tennessee, Knoxville, Department of Computer Science Technical Report CS-96-331.

MacLennan, B. J. (2008). Consciousness: Natural and artificial. *Synthesis Philosophica*, 22(2), 401–433.

MacLennan, B. J. (2009). Robots react but can they feel? A protophenomenological analysis. In: J. Vallverdú & D. Casacuberta (Eds.), *Handbook of Research on Synthetic Emotions and Sociable Robotics: New Applications in Affective Computing and Artificial Intelligence* (pp. 133–153), Hershey, NJ: IGI Global. doi:10.4018/978-1-60566-354-8.ch008

MacLennan, B. J. (2010). Protophenomena: The elements of consciousness and their relation to the brain. In A. Batthyány, A. Elitzur, & D. Constant (Eds.), *Irreducibly Conscious: Selected Papers on Consciousness* (pp. 189–214), Heidelberg & New York, NY: Universitätsverlag Winter.

MacLennan, B. J. (2011). Protophenomena and their physical correlates. *Journal of Cosmology*, 14, 4516–4525. Reprinted in R. Penrose, & S. Hameroff (Eds.), *Consciousness and the universe: Quantum physics, evolution, brain & mind* (pp. 228–239). Cambridge, MA: Cosmology Science Publishers.

MacLennan, B. J. (2014). Ethical Treatment of Robots and the Hard Problem of Robot Emotions. *International Journal of Synthetic Emotions*, 5(1), 10–18. doi:10.4018/ijse.2014010102

MacLennan, B. J. (2015). Living science: Science as an activity of living beings. *Progress in Biophysics and Molecular Biology*, 119, 410–419.

McCall, R. J. (1983). *Phenomenological psychology: An introduction. With a glossary of some key Heideggerian terms*. Madison, WI: University of Wisconsin Press.

Melzack, R., & Casey, K. L. (1968). Sensory, motivational and central control determinants of chronic pain: A new conceptual model. In Kenshalo, D. R. (Ed.), *The skin senses: Proceedings of the first International Symposium on the Skin Senses, held at the Florida State University in Tallahassee, Florida* (ch. 20). Springfield: Charles C. Thomas.

Morsella, E., Godwin, C. A., Jantz, T. K., Krieger, S. C., and Gazzaley, A. (2015). Homing in on consciousness in the nervous system: An action-based synthesis. *Behavioral and Brain Sciences*, 106 pages. Published online 22 June 2015. Available on CJO2015: doi: 10.1017/S0140525X15000643

Pettit, D. L., Wang, S. S., Gee, K. R., & Augustine, G. J. (1997). Chemical two-photon uncaging: A novel approach to mapping glutamate receptors. *Neuron*, 19(3), 465–471. doi:10.1016/S0896-6273(00)80361-X PMID:9331338

Plutchik, R. (2000). *Emotions in the practice of psychotherapy: Clinical implications of affect theories*. New York: American Psychological Association.

Plutchik, R. (2003). *Emotions and life: Perspectives from psychology, biology, and evolution*. New York: American Psychological Association.

Plutchik, R., & Conte, H. R. (Eds.). (1997). *Circumplex models of personality and emotions*. Washington, DC: American Psychological Association.

Prinz, J. (2006). *Gut reactions: A perceptual theory of emotion*. Oxford: Oxford University Press. doi:10.1093/acpro f:oso/9780195309362.001.0001

Rolls, E. T. (2005). *Emotion explained*. Oxford, UK: Oxford University Press.

Rolls, E. T. (2006). *Brain mechanisms of emotion and decision-making*. International Congress Series, 1291, 3–13. Amsterdam, Netherlands: Elsevier.

Rolls, E. T. (2007). A neurobiological approach to emotional intelligence. In G. Matthews, M. Zeidner & R. D. Roberts (Eds.), *The science of emotional intelligence* (pp. 72–100). Oxford, UK: Oxford University Press.

Rudrauf, D., Lutz, A., Cosmelli, D., Lachaux, J.-L., & Le Van Quyen, M. (2003). From autopoiesis to neurophenomenology: Francisco Varela's exploration of the biophysics of being. *Biological Research*, 36(1), 27–65.

Service, R. F. (2013). The cyborg era begins: Advances in flexible electronics now makes it possible to integrate circuits with tissues. *Science*, 340(6137), 1162–1165. doi:10.1126/science.340.6137.1162 PMID:23744924

Simon, H. (1977). What computers mean for man and society. *Science*, 195(4283), 1186–1191.

Strawson, G. (1994). *Mental reality*. Cambridge, MA: MIT Press.

Strawson, G. (2006). Realistic monism. In A. Freeman (Ed.), *Consciousness and its place in nature: Does physicalism entail panpsychism?* (pp. 3–31). Charlottesville, VA: Imprint Academic.

Sur, M. (2004). Rewiring cortex: Cross-modal plasticity and its implications for cortical development and function. In G.A. Calvert, C. Spence & B.E. Stein (Eds.), *Handbook of multisensory processing* (pp. 681–694). Cambridge, MA: MIT Press.

Varela, F. J. (1996). Neurophenomenology: A methodological remedy to the hard problem. *Journal of Consciousness Studies*, 3, 330–350. Reprinted in: J. Shear (Ed.), *Explaining consciousness: The hard problem of consciousness* (pp. 337–358). Cambridge, MA: MIT Press, 1997.

Weisberg, J. (2012). The hard problem of consciousness. *The Internet Encyclopedia of Philosophy*, ISSN 2161-0002. Retrieved January 4, 2016 from http://www.iep.utm.edu/

Weiskrantz, L., Warrington, E. K., Sanders, M. D., & Marshall, J. (1974). Visual capacity in the hemianopic field following a restricted occipital ablation. *Brain*, 97, 709–28.

KEY TERMS AND DEFINITIONS

Action Potential: A neural impulse, an electrical signal generated by a self-reinforcing electrochemical process, by which neurons communicate with other neurons, generally conveying information by their firing rate (rate of action potential generation).

Activity Site: A physical structure (in an animal or robot) corresponding to a protophenomenon (q.v.), so that the protophenomenal intensity (q.v.) is correlated with physical activity at the activity site.

Intensity, Protophenomenal: The degree of a protophenomenon’s presence in the conscious state, which is correlated with physical activity at an activity site (q.v.).

Neurophenomenology: An approach to neuropsychology or, more broadly, to studying the human mind, which combines neuroscientific techniques with phenomenology (q.v.), with investigations in each domain informing and reinforcing those in the other.

Phenomenon: Anything definite that arises in consciousness, including perceptions, sensations, feelings, recollections, dreams, hallucinations, desires, intentions, and imagination.

Phenomenology: The study of the structure of conscious experience by means of systematic introspection and analysis. There are many alternative approaches to phenomenology.

Protophenomenal Analysis: An analysis of some domain of conscious phenomena in terms of protophenomena (q.v.) and their neural correlates.

Protophenomena: Hypothesized smallest constituents of a conscious state, its elementary degrees of freedom.

Qualia: The fundamental phenomenal qualities associated with a conscious experience, for example, the experience of blueness when observing or imagining a blue object.