Experiencing: A Jamesian Approach

Simona Ginsburg[1] and Eva Jablonka[2]

Abstract

This paper suggests an approach to consciousness that focuses on the evolutionary transition from pre-conscious animals to the simplest types of conscious (experiencing) animals. Our argument is that experiencing originated with the evolution of associative learning, and that one of the major functions of experiencing was what William James called 'fighting for ends': endowing animals with motivation. We propose that the sensory states generated during associative learning act as internal guides and selectors of new neural relations, new behaviors, and new ends, leading to the unitary, subjective and intentional internal dynamic states that we recognize as experiencing.

[1] The Open University of Israel
[2] Tel Aviv University, Israel

1. Introduction

Several complementary approaches have been used to try to uncover the neural processes underlying consciousness, the fundamental ability to experience (e.g. feeling hungry). These include mapping differences between the neural dynamics of subjects during the wakeful state and during deep, non-dream sleep; during voluntary versus non-voluntary acts; during implicit versus explicit learning; in neurologically impaired subjects, who have lost the awareness of certain types of experiences (as in blindsight) and in normal subjects; in people with special types of awareness (synesthesia) and normal individuals (reviewed in Chalmers, 2004). Although these approaches have provided important insights into consciousness, they typically deal with very complex animals – almost invariably vertebrates. Consciousness in these animals has evolved for a long time, and its properties are therefore highly derived.

The approach we take focuses on the *evolutionary transition* to experiencing, from pre-conscious to conscious animals. This has the advantage that the systems analyzed are relatively simple, and hence derived dissociations and integrations that occurred at later evolutionary stages do not mask the fundamental properties of the process of experiencing. In previous papers we have argued that associative learning in animals with a highly interconnected, integrating central nervous system (CNS) is an identifying criterion for consciousness, and that it appeared early in metazoan evolution (Ginsburg & Jablonka 2007a,b). Here we extend our arguments by taking and developing a Jamesian view of consciousness, and show how a new type of telos – individual motivation – emerged with experiencing.

2. Precursors and gray areas: limited experiencing and limited learning

The transition-based approach that we describe seeks to show how the simplest possible experiencing organisms evolved from those that did not experience. Like all biologists, we believe that the basis of experiencing is the rich interconnectivity that enables neural communication and integration. In all neural animals, neurons are organized as a system; there are no animals that have several, completely unconnected, nerve-nets. Neural connectivity is not enough, however: the transition from sensory processing and integration, however rich, into experiencing, however limited, requires something more. Philosophically, there is a categorical leap here,

similar to the leap between inert and living matter. Although we cannot pinpoint what turns sensory processing into experiencing, we believe that if we recognize that this leap did occur during evolution, and can identify when it occurred, we can explore the processes and the organization involved in the transition. This will take us a long way towards a better understanding of experiencing.

The simplest type of nervous system that exists in extant animals is that of medusas of the phylum Cnidaria, and it is plausible that its organization is similar to that of the earliest (extinct) animals endowed with a nervous system (Nielsen, 2008). The Cnidaria are seen today as a sister group of the triplobalstic bilaterians, the bliaterians that have true mesoderm and that include the acoelomorph flatworms, the group from which the protostomes and deuterostomes had evolved (Baguňà and Riutort 2004, Erwin 2006). The cnidarians' nerve net is seen as an ancient trait, sharing basic characteristics with the diffuse nervous system of the acoelomorphs (Baguňà et al 2008). Despite its relative simplicity, the nervous system of cnidarians already has the organizational building blocks that are the basis for experiencing (Ginsburg & Jablonka, 2007a, references therein). These are:

1. <u>An interconnected nervous system that enables sensory and motor coordination, and the integration of information coming from multiple sources.</u> The nervous system of cnidarians is a diffuse net with interconnected neurons, no brain and little centralization (although there is some concentration of nerves in the apical organ in the larvae and around the mouth in adults). Stimulating one nerve cell triggers electrical signals in the entire net. A chain-reaction of neural activity is set up because internal and external sensors trigger motor activity, the motor actions trigger other sensory neurons, and so on. We call the global activity in an interconnected nervous system, at any moment, *overall sensation*. The overall sensation includes neural activities that result from normal, homeostatic maintenance, as well as those occurring as the animal responds to various conditions that threaten its survival or are necessary for its reproductive activity. As a metaphor, the overall sensation can be regarded as white noise – a functionless by-product of a sensory-motor system that dynamically processes electrical and chemical signals. The activity of pacemakers that provide ongoing rhythmical internal stimulation, which can be found in all neural animals including cnidarians (Mackie & Meech 2000), probably imparts a certain flavor to the

3

animal's overall sensation. The overall sensation is as yet not experiencing, but it is the evolutionary raw material from which it will emerge.

2. <u>Reflexes.</u> As well as 'directed' reflexes, in which an adaptive, highly specific, motor response follows a specific type of stimulation (such as the withdrawal response to touch), medusas have 'non-directed' reflexes. With these, the motor responses to a particular sensory stimulation are exploratory; for example, lack of food stimulates random searching behavior. A reflex act can follow a single stimulus, such as touch, but here we want to focus on reflexes that occur when there are persistent departures from homeostasis, such as lack of food or salts, tissue damage, or hormonal excitation. With this type of reflex, internal sensors are continuously activated, signaling an out-of-equilibrium state, until homeostasis is restored. This is a process that can take time.

Non-directed reflex motor activity is an adaptive response: an animal seeks to change its environment either because its present conditions are detrimental and it cannot restore homeostasis, or to initiate reproduction. The animal's mal-adaptive or internally excited (non-homeostatic) condition can be thought of as a *repulsor state – an internal physiological state which the animal tries to alter.* Usually, in order to alter such an internal state, the animal changes it position in space, i.e., it moves. In very simple animals there is no guide for the animal's action: they move at random. Although such random movements are beneficial, because they change the animal's location, any mechanism that *guides* the animal so that it reaches a more suitable place and achieves an adaptive state more readily is obviously advantageous. Neural organisms, including cnidarians, have such mechanisms.

3. <u>Exploration and selective stabilization processes within the nervous system that guide behavior.</u> All selective stabilization processes are based on the generation of a large set of local variations from which only a small subset is eventually stabilized and manifested. The eventual output depends on the initial conditions and the number of possible points around which the system can be stably organized. These points, known as *attractors, are the set of states towards which the system tends to proceed, regardless of the conditions from which it started.* There are high level and low level attractors. The high level attractors on which we focus here are the internal

homeostatic states that the animal tries to attain through its behavior, and to maintain once reached; they depend on lower level attractors and the exploration and selective stabilization processes underlying them. The way to a major attractor, such as satiation, can be strewn with several "stepping stones" or lower level attractors. These lower level attractors are intermediate equilibrium states (or 'checkpoints') that make the exploration more reliable.

There are many examples of exploration and selective stabilization in the nervous system. During embryonic differentiation, many neurons die unless their fibers are stabilized by accidentally-met chemicals known as growth factors. In addition, synapses – the connection points between neurons – are at first formed in vast numbers, but many are lost: they survive according to the rule 'neurons that fire together, wire together', and if two neurons are not active together, the synapse disappears.

Of particular importance to our arguments is the type of exploration and selective stabilization that leads to open-ended adaptive behavior. Even in animals like cnidarians, which do not learn associatively, there are mechanisms that transiently reinforce the activity of the neural circuits that promote the *repeated* stimulation of sensors (such as food sensors) that indicate fitness enhancing conditions (such as consuming food). Similarly, the inhibition of neural circuits that lead to *repeated* stimulation of sensors indicating fitness-reducing conditions is reinforced. We assume that the neural trajectories of stabilized excitation and inhibition are built on the existing reflex pathways embedded within the randomly generated neural connections formed developmentally. However, by recruiting additional, interconnected, *newly-used* parts of the nerve net, the networks formed went beyond the reflex pathways.

4. <u>Simple learning mechanisms that can modify reflex behavior.</u> Learning requires the ability to memorize a stimulus-response relation, and to recall it. Habituation and sensitization, the elementary learning mechanisms of cnidarians, are modulations of directed reflexes in which the synaptic connections among neurons are strengthened or weakened. With habituation, repeated stimuli are ignored or responded to in a diminished manner. For example, when a stimulus, such as a light touch, causes no deviation from homeostasis (or when the reflex response leads to a stronger repulsor

5

state than that presented originally), it is ignored. Conversely, with sensitization the magnitude of a response to a repeated stimulus is increased, or its threshold is lowered. For example, withdrawal from a predator is more rapid, and movement towards a food source is faster.

3. The basis of limited experiencing: persistent, integrated sensory stimulation

Assume that continuous and specific sensory stimulation, such as that resulting from lack of nutrition, lasts for some time. This persistence will impart a consistent sensory effect on the background of overall sensation. It is a kind of sensory signature, which is specific to this ongoing signaling of a non-homeostatic state. We suggest that this signature is the basis of experiencing. We call such specific, persistent and integrated states *limited experiencing* states (Ginsburg and Jablonka 2007a). They are limited because their number is relatively small, and because the sensory signature that defines them has as yet no function, so they cannot evolve by natural selection.

Herbert Spencer, among others, thought that the persistence of altered, yet integrated, neural activity is absolutely necessary for an animal to become conscious (Spencer, 1855, pp. 591-2). The importance of a minimal length of time for experiencing to occur was stressed by William James, who following Clay, called it the 'specious present' (James, 1890, volume I, p. 609). Later scientists also highlighted temporal persistence as a necessary aspect of experiencing. Homer Smith (1959) regarded the persistence of integrated neural events that form an enduring image (lasting seconds and even minutes) as the essence of consciousness. Humphrey (1992, 2006) called this temporal persistence the 'extended present', and Edelman (1989) referred to it as the 'remembered present'.

Behavioral exploration involves synchronous stimulation of groups of sensors and effectors, some of which are transiently stabilized. For example, the sensors related to food sensing are continuously stimulated, and so are some of the effectors that lead to movement towards the food source. This means that some sensory-motor neural activities persist and are super-imposed on the white noise of overall sensation. They can therefore be considered as additional sources of limited experiencing. There is a difference between a persistent stimulation that does not involve selective stabilization (for example, when there is continuous food deprivation but no food

source is encountered) and one that does. In the first case, only neural stimulations caused by the repulsor state are the source of limited experiencing. In the second case, food seeking involves movement along a food gradient, so some of the neural trajectories associated with the attractor state may also become activated (for example, because the initially sparse food molecules trigger activities that are involved in reaching the homeostatic state of food-adequacy).

The sensory signatures of persistent stimulations that signal a fitness-relevant change in the internal state of the organism (one that requires homeostasis-restoring behavior) leads to what may be seen as the simplest types of limited experiencing. They are limited not only because they correspond to the few directed or non-directed reflex actions of the animal and, because they are functionless, they cannot be the basis of evolution by natural selection, but also because the integration of information in the nervous system is still limited. These types of limited experiencing seem to correspond to states that Denton (2006) called 'primordial emotions': 'emotions' that are associated with departures from homeostasis and that mostly depend on continuous and synchronous activity of multiple interoceptors. However, for these internal states to have any functional significance, they must lead to appropriate responses to the environment, which depend on the stimulation of exteroceptors. Crucially, from our point of view, these sensory signatures (limited experiencing), which correspond to 'primordial emotions', are at this stage no more than by-products of the global, consistent and specific neural activities imposed on overall sensation. As we argue in the next sections, with the evolution of associative learning, limited experiencing acquired functions: it enabled discrimination on the basis of partial inputs and endowed animals with drives and motivation.

4. Heading towards experiencing: the evolution of associative learning

We argue that the great transition that generated a new telos, individual motivation, was a result of the evolution of associative learning. This evolution involved the ability to retain memory traces of previously selectively stabilized connections even when the stimulus was no longer present. We believe that the acquisition of the ability to form such memory traces involved the recruitment, elaboration and sophistication of the memory systems underlying habituation and sensitization. When fully-fledged associative learning emerged, the limited learning based on sensitization and

habituation became more flexible, and evolutionarily extensive. This brought about the transition from limited to unlimited experiencing: specific overall sensory states became fully-fledged feelings. By 'fully-fledged feelings' we mean global sensory states that act as motivators of individual behavior.

Associative learning is a flexible type of learning which requires remembering and evaluating new associations between sensory stimuli, or between sensory stimuli and responses (Razran, 1971). It is customary to distinguish between two major types of associative learning – classical (Pavlovian) conditioning and instrumental or operant conditioning. Classical conditioning is a modification of behavior in which a new 'neutral' stimulus is paired with a stimulus that already elicits a particular response, either because it is innate or because it was learnt at an earlier stage. An organism exposed repeatedly to pairs of the 'neutral' and original stimuli will eventually respond to the neutral stimulus alone. In terms of repulsors and attractors, the repulsor is the internal state (e.g. 'hunger'), which leads to motor activities involved in the search of food and the preparation for its consumption, until the attractor, the internal homeostatic state (e.g. 'satiation') is reached. Instrumental conditioning is a form of associative learning in which the *actions* of the organism are reinforced by their consequences. As in the previous case, the repulsor may be a state of 'hunger' and the attractor the internal homeostatic state of 'satiety', but this time reaching the attractor is associated with new motor acts, rather than newly associated sensory stimuli.

Associative learning involves the formation of *new* connections in the neural circuits in which sensors and effectors are embedded. Crucially, neural traces of a new relation are maintained even in the absence of the stimulus, and it is recalled (the trace is activated) when the stimulus appears again. In theory, the number of associations is large even when the number of sensors and effectors is modest. In practice, there are constraints on the number of neural connections possible, but the links formed are not pre-determined by past evolutionary history, and are selected from a very large repertoire of developmentally generated connections.

All animals for which associative learning has been reported are protostomes and deuterostomes: bilaterians characterized by anterior brain ganglia and nerve cords running along the body. In contrast, studies of learning in cnidarians and ctenophores,

who have a diffuse nervous system and no brain, failed to find definite instances of associative learning (Corning et al 1973). The cnidarians are not, however, the direct ancestors of the protostomes and deuterostomesm but rather their sister group. It is now believed that the acelopmorphs, who like the cnidarians and ctenophores have a diffuse nervous system but do show some clustering of nerve cells at the anterior end of the animal forming a cerebral commissure, are the immediate ancestors of the protostomes and deuterostomes (Baguňà 2008). The learning ability of acoelomorphs has, unfortunately, not been studied, but their small size, the simple organization of their diffuse nervous system, and the lack of proper brain make it unlikely that their learning ability is more advanced than that of the cnidarians.

It is plausible that a well defined brain and nerve cords are necessary for the evolution of associative learning, but it is unlikely that it was sufficient; brain memory mechanisms enabling the stabilization of transient neural links following exploration and selective stabilization must also have been involved. We propose that more reliable and flexible memory mechanisms evolved following an increase in size that occurred as a result of the increased oxygenation of the sea during the later Ediacaran period, ~548 million years ago (Fike et al., 2006). Although the amount of oxygen required for some activities, such as slow burrowing in mud, is low (Budd 2008), high oxygen levels have been shown to be important for sustaining complex ecology (Catling et al. 2005) and for the types of activities, such as swimming, assumed to occur in the relatively large Cambrian animals. An ancestral small acoelomorph that already possessed anterior clusters of nerves and muscles had the potential to control its movements when it grew in size, so selection for coordinated movements would have been important. Crucially, increased oxygenation is important for neural tissues, which are metabolically expensive, so their growth and maintenance was probably increased as the level of oceanic oxygen rose, extending neural connectivity and enabling the formation of new connections. The benefits of learning increased when the life span (which is likely to have been correlated with size) increased, for animals lived long enough for past events to recur and for their memorization to be worthwhile.

The new memory mechanisms probably included the physical formation of new micro-anatomical connections and the thickening and modifications of synaptic

structures. Neural cellular mechanisms, such as back-propagating action potentials, long-term-potentiation (LTP), and stabilizing patterns of re-entry (back and forth signaling between large ensembles of neurons that lead to the temporal persistence which is necessary for binding and for experiencing during recall) must have also been involved (Edelman & Tononi 2000; Edelman, 2003).

The long-term stabilization of new associations, which requires such new types of memory mechanisms, led to the evolution of associative learning (Ginsburg & Jablonka, 2007b). The nature of the molecular changes that may have been involved is as yet unclear. However, recently Emes et al. (2008) compared the genomics and proteomics of the postsynaptic density and the membrane-associated guanyl kinase-associated signaling complexes that underlie memory and learning, and found interesting differences between invertebrates and vertebrates; this suggests that a comparison of synaptic proteins of associatively-learning protostomes and deuterostomes with that of acoelomorphs, cnidarians and ctenophores, may be fruitful in uncovering the molecular correlates of associative learning.

5. The teleological function of experiencing: extending the Jamesian approach

To explain how, through the evolution of associative learning, specific overall sensations became motivating experiences, our starting point will be William James' (1890) scheme for producing feelings. His famous example of meeting a bear, though only briefly described in his major book (volume II, pp. 450-1), has become paradigmatic. James states that on meeting a bear for the first time, an innate reflex or instinct is activated, and this instinct is 'blind', it has no 'quintessence'; the stimulus is an unstructured 'something', a trigger that activates a specific motor reflex response of fleeing. The end, too, according to James – the shelter reached following escaping – is unembodied; but if this first meeting with the bear lasts for a minimal amount of time, *a feeling is formed*, due to feedback sent from the animal's body to its brain during the motor response. Note the serial chain in James' scheme (Figure 1):
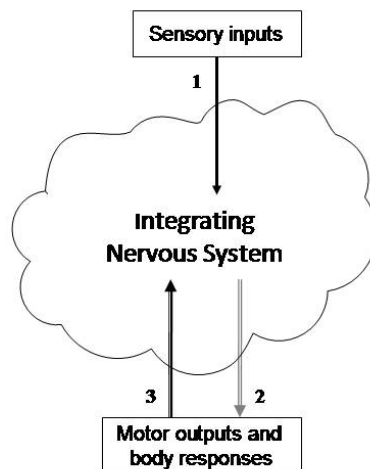
Figure 1: James' basic scheme

The stimulus is received in the nervous system (1), and activates a reflex motor response (2); signals from sensors that are activated by motor acts are fed back to the brain (3). These signals, which may include the sensory inputs from elevated heartbeats and perspiration that accompany the fleeing response, are integrated in the central nervous system, and comprise the feeling of fear. Hence, fear does not precede flight, but results from all the integrated (bound) bodily, persistent sensory-motor responses associated with fleeing. A famous quote makes this point very clear: '…we feel sorry because we cry, angry because we strike, afraid because we tremble, and not that we cry, strike or tremble, because we are sorry, angry, or fearful, as the case may be' (volume II, p. 450).

On further encounters with a bear, the shapeless stimulus does become embodied, according to James. If the threatened animal memorized some of the inputs related to the triggering of the reflex, the route taken towards safety, and the end (safety), then the future responses cease to be blind, and so do the ends; because bodily activities that accompanied the first flight are remembered, and so is the context of the first encounter with the originally shapeless trigger-bear (e.g. the bear's humming voice)

11

as well as the shelter found, everything becomes embodied. As James put it, 'it is obvious that *every instinctive act, in an animal with memory, must cease to be 'blind' after once repeated*, and must be accompanied with foresight of its 'end' just so far as that end may have fallen under the animal's cognizance.' (volume II, p. 390, James' italics). The function of consciousness is thus to achieve ends, ontogenetically: 'Every actually existing consciousness seems to itself at any rate to be a *fighter for ends*, of which many, but for its presence, would not be ends at all.' (volume I, p. 141, James' italics.)

James did not offer a mechanism that explains how 'fighting for ends' actually occurs. We suggest a mechanism based on associative learning, and our scheme (Figure 2) may be viewed as a reconstruction and extension of James' views. We see the first encounter with a stimulus as James did. But during the subsequent encounters, in addition to activating the serial chain 1-2-3, we suggest that *in parallel*, and concomitantly, memory traces of past encounters are activated as well (the broken arrows in figure 2). These are memories of previous similar inputs, motor acts, contexts, and results, i.e. the stimuli related to the repulsor state, to the attractors reached by the animal, and the pathways that led to them. It is important to realize that activated memory traces of past responses and past ends, which could not exist in the first encounter, *are part of the input into the nervous system in further encounters with the external triggering stimulus.* The memorized inputs are internal, so the animal responds to both external inputs and to the activated memorized internal neural inputs.
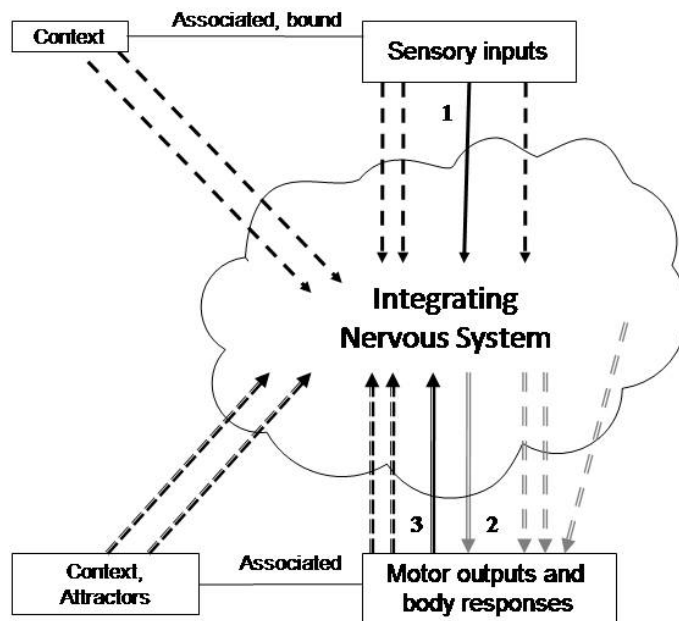
Figure 2: Extending the Jamesian scheme

At each moment, the sum-total of neural activity in the animal's nervous system, due to persistent stimuli and the activation of memory traces which have a minimal temporal, present-extending duration, is an overall sensory state – a dynamic state with a specific sensory flavor or signature, an experiencing. An animal capable of learning associatively recalls the past-encountered, learnt, stimulus-response relations. The neural consequences of learning are integrated with the original sensory state, and are remembered. When the behavior is re-triggered, the overall sensory state, which includes the activated internal traces of the embodied trigger for action (the bear in its embodied glory), the activated internal traces of the action taken, and the activated internal traces of the attractor-related stimuli (the actual shelter) can now direct exploratory behavior upon encountering the same or similar stimuli again. The goal-associated activated traces that are included in the experiencing, as well as some of the ways of reaching the end, function as pointers to this goal.

The example of escaping a bear and the emotion of fear that it elicits are already highly evolved reactions. However, associative learning in an invertebrate such as a crab follows the same basic principles. The bound elements perceived and remembered may be very few and very restricted, but nevertheless the synchronous activation of past associations, such as those associated with food, is inherently motivating. Let us look at the reaction of a crab to lack of food. The notion of food does not at first exist. It is only after the animal has learnt that particular things in the world are associated with relief of nutrition-lack that it can identify them as food and they can act as mediating attractors, as ends. These ends did not and could not exist before learning. They are learning-constructed, embodied goals that lead to corresponding internal attractors. Associative learning also enables the animal to embody the repulsor-triggering input: the triggering input may be very partial – for example, an associatively learnt cue such as a certain type of vibration, which initially did not trigger the food-searching behavior. This, however, is enormously beneficial, since the detection and response to stimuli associated with a repulsor state can often be life saving.

There are therefore three major advantages to associative learning. First, the binding of stimuli, which at the previous stage of evolution (before associative learning had evolved) merely accompanied experiencing, becomes advantageous, because it makes it possible to distinguish between complex (combined) stimuli. Second, learning-dependent experiencing allows recognition and discrimination on the basis of *partial cues*: for the hungry animal contingent associations (e.g. vibrations) may be recognized and elicit an adaptive response, food-seeking. Third, since for the food-deprived but already experienced crab, food had become embodied, it gives the animal clues as to *what to do*, since some of the activated traces are associated with successful navigation towards the attractor-related stimuli (food, shelter and their contexts). The animal can now make an *educated guess*, based on its past experience. All these repulsor and attractor related stimuli are partially overlapping, dynamically changing aspects of the overall sensory-motor state of the animal, driving its activities.

All neural connections that were memorized and become activated on a second encounter become part of the ceaselessly active nervous system, a part of overall

14

sensation. They confer on the nervous system a particular distinct and persistent sensory signature which is different from those previously described because now these sensory states are the consequence of the animal's learning history. We have called any activated, distinct, history-dependent, persistent overall sensory state a *categorizing sensory state (CSS)* (Ginsburg and Jablonka 2007a,b). At the organism level, a CSS is a global dynamic sensory neurophysiological state – either a repulsor or an actively maintained attractor state. We call this sensory state 'categorizing' for two reasons. First, because the inputs that elicit it activate memory traces of other inputs of *the same type*, for example, inputs and memory traces related to tissue-damage. Second, it is 'categorizing' because the inputs and traces will determine what type of response will occur, as memory traces of the motor responses to the inputs and the stimuli associated with the attainment of the attractor (relief from tissue damage) are activated.

An essential aspect of a CSS is its dynamic nature. Suppose the animal is in a particular global sensory state, that of food deprivation, and begins exploring. Among the activated parts of the network are memory traces of the impact of food (which is related to the attractor state) and the route taken to reach it, but most dominant at first are the memory traces related to triggering the repulsor state. The various activated traces are built on innate connections but also, and crucially, on previously formed associations. In Figure 3 the network associated with the dominant repulsor-state is gray. The dotted network in the same brain represents other neurons that are activated in the attractor state in which the animal settles when satiated. Because initially the animal is hungry, the CSS reflecting this is not fitness-promoting (the CSS is a repulsor), and the animal will tend to change it; the gray CSS thus acts as a motivator for actions that will lead the animal to change this CSS towards the dotted one. Exploration is to some extent guided by the activation of memory traces that were formed during past encounters that led to a sensory state that corresponded to satiation. As a result of these exploratory motor actions, food may actually be found.
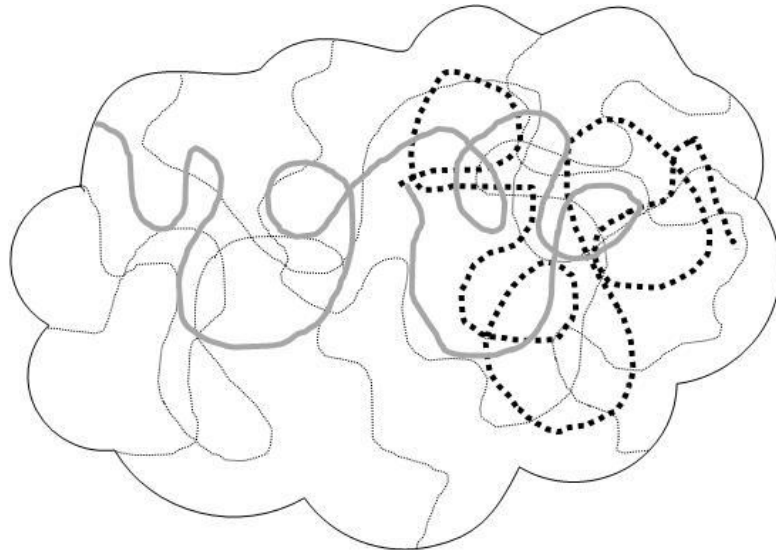
Figure 3: Two CSSs

The gray network denotes an ensemble of brain neurons that are activated when the animal is in a particular repulsor state (hunger), and the dotted lines denotes the neuron networks activated when the animal reaches an attractor state (satiation). The thinner lines are other neural networks that are not components of either of these two states. The CSSs are specific to a particular occasion: on other occasions, the CSSs will be somewhat different, although there will be family resemblances between them.

Figure 4 illustrates the dynamic movement from the gray to dotted CSS. At the beginning (left), most of the gray network is activated and just a few parts of the dotted CSS are active; activated neurons in the gray CSS might include glucose sensors, memory traces of past encounters of food items, the context in which they occurred, etc. In the middle part, some motor pathways are activated as the animal explores its surrounding: an associated smell is encountered, and the animal moves towards it. At this time both CSSs are activated. Finally, food is caught and consumed, and when the attractor – satiation – is reached, the dotted CSS is much more activated (right).
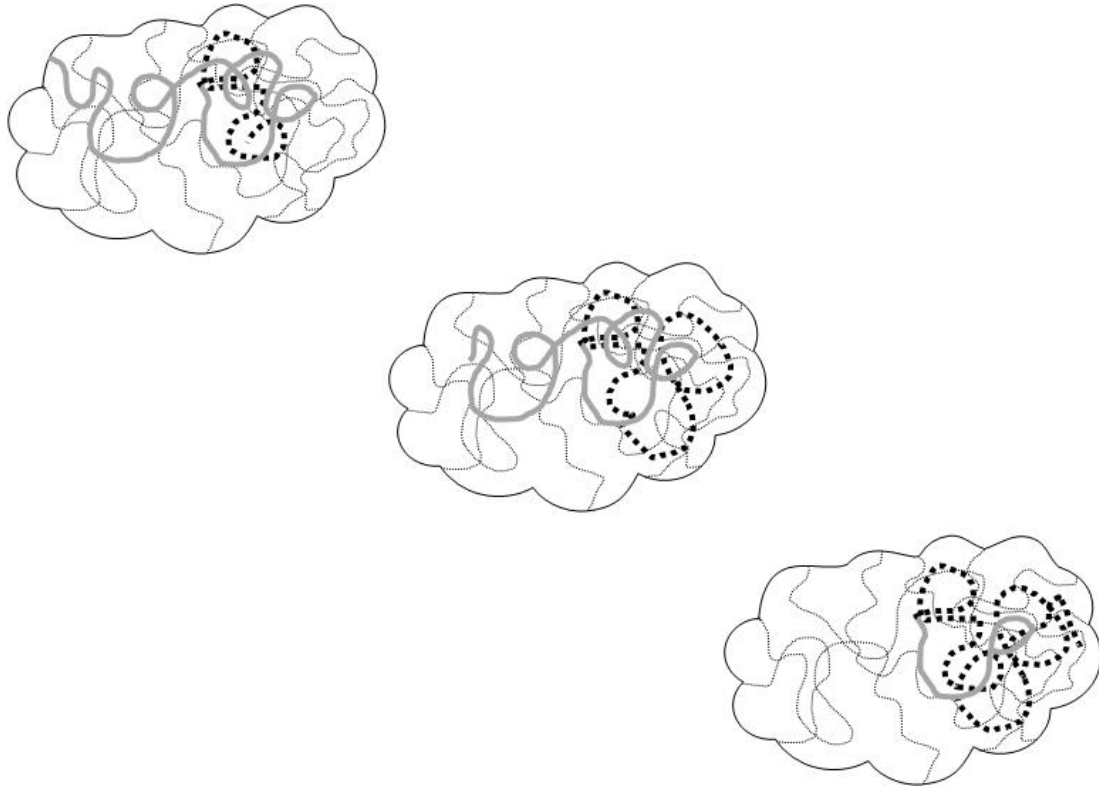
Figure 4: Moving from repulsor to attractor

According to this Jamesian scheme, every CSS is built around a basic, innately given reflex trajectory. These reflexes are related to the fundamental physiological functions of reproduction and self maintenance, such as the acquisition and consumption of food, and the maintenance of tissue integrity and an adequate salt balance (Denton 2006). The internal state of homeostasis or departure from it is signaled by interoceptors, and exteroceptors are used to respond adaptively to the environment in order to restore homeostasis. However, once associative learning evolved, the fixed reflex trajectories can be thought of as a neurophysiological scaffold, which can become redundant after its initial activation, so that a particular CSS in the adult may be rather different from the original, fixed-reflex. In the same individual at different stages of its life, the CSS leading to food-seeking behavior, for example, may be different in detail, because the learning history and memory of the encounters with the food stimuli are different. The patterns of connections among neurons that are activated during a given type of response have family resemblances, but are not identical; they have many connections in common, but no single connection is obligatory. The effects of a new stimulus or response can become part of several different CSSs; they can belong to the CSS that elicits food-seeking, or withdrawal, or

17

partner-seeking. Although CSSs have great flexibility, each is constrained to a certain domain of the animal's actions and is robust. The number of CSSs that any one animal at any particular stage of life may have is limited, but the number of *potential*, newly learnt associations within a CSS is large, because many new links between stimuli and responses could belong to the same CSS.

Once CSSs are formed, the neural effects of a single incoming stimulus are produced in the context of other, pre-existing and previously formed neural trajectories. The integrated learning-based internal state, rather than the effect of a single stimulus, becomes the cause of the behavior. The results of the external stimulus become just a part of the system, contingently related to the direct cause – the CSS. The particular, unitary and global nature of the CSS, its particular quintessence, guides the animal to the right realm of action, while its specific history-dependent features give it an individual flavor and a more or less specific action-direction. Hence, *once CSSs evolved, they became central organizing causes and navigators for the animal's goal-oriented actions.* An overall sensory state became a state *for* the animal, not just a state *of* the animal. It became part of a new teleological system. Motives had come into being.

Let us reiterate: the function of the overall sensory state, which is a particular CSS with its activated traces and with its specific yet global sensory effect, is to inform the animal about its present deviation from the attractor state, and guide it towards reaching it by directing adaptive behavior that is based on past history. Whereas an animal with no associative learning cannot remember associated cues and has to depend on the original external cues to find the attractor (e.g. depend on the actual food traces that enable it to follow a food gradient, rather than vibrations associated with food), an animal with associative memory has the great advantage that it can respond to associated cues pointing to the attractor state. The CSS is therefore both an internal discriminator and a motivator: it evaluates inputs on the basis of the animal's previous learning history, which includes past triggers of exploration, the route taken, *and the goal attained*. Even in the absence of external attractor-related cues, the initial exploration is partially guided by the attractor-related *internal* memory-traces. The animal is therefore endowed with a *'remembered future',* not just a remembered present. Thus, through the CSS, the animal 'informs' itself what to do, in what

manner, and towards which end. Such an animal has an enormous adaptive advantage over one that does not have such future-oriented memories, such internal 'guidance'.

Our scheme thus departs somewhat from James': although we agree with his claim that experiencing emanates from specific sensory-motor physiological correlates, these correlates are different during the first and the 'second' encounter with the stimulus. The 'second' time need not be the chronologically second encounter – it may sometimes be the third or fourth; it is defined by whether or not recall is possible, not by its chronological position. On this 'second' encounter, there is already memory of embodied inputs, activities, and attractor-related stimuli (the remembered future). Triggering experiencing includes the activation of the memory traces of some of these, as well as the initiating input (although *memory alone* can drive an animal to action). The inputs and the activation of all these memory traces lead to a dynamic overall sensory state that overlaps, but is not identical with that of the first experiencing. The animal is afraid *before* it actually flees and its fleeing is to some extent informed, because the memorized fleeing-related neural trajectories are stimulated (with some, but initially not all, of their motor correlates). Moreover, this memory-based experiencing drives exploration – away from the repulsor state, towards the attractor state. All these generate the overall, unitary sensory state we call fear. Fear now precedes and causes fleeing towards a refuge, and may even precede trembling. This fear-CSS cannot, ever, be repeated exactly, but all fear CSSs have a family resemblance, based on and related through the original scaffolding reflex. Similarly, following learning, hunger is *for embodied food*, and anything associated with it (its smell, the route to it, the coincident vibration of the water) will guide the exploration process.

Our view of consciousness agrees with that of James and most other neurobiologists and philosophers in regarding consciousness as a process, as experiencing, which is a system-level feature of the animal's integrated physiology, not of the brain alone (Damasio 1994, Bennett and Hacker 2003). Our suggestion that CSSs are the stuff of experiencing is compatible with the properties that many philosophers attribute to consciousness. Searle (2004), for example, suggested that consciousness is a system-level biological state of the brain, caused by neurobiological processes, and that conscious states function causally in directing behavior. In addition, Searle

19

characterizes consciousness as unified, intentional, qualitative and subjective. Edelman (2003) gives a more exhaustive, but essentially very similar list of characteristics. CSSs fit all of these criteria: they are, by definition, unitary; they are always part of a global and integrated physiological system of the whole organism. They are also intentional, in the sense that they are 'about' events and relations, as well as in the sense that they express the individual's needs and intentions. As we have argued, one of their important functions is as internal motivational and reward systems. CSSs are also qualitative: since each CSS is a specific configuration of neural activity operating as part of a globally active system (a body), it has a distinct, yet whole-organism character – a domain-specific individual quality, which cannot be reduced to an aggregation of stimuli. Finally, CSSs are subjective: since they are internal and dynamic activity-states that are the result of individual developmental and learning processes, they are idiosyncratic and unique, depending on the reverberating, neural activities within the whole animal, which are fully accessible only to itself.

Drive is defined in the *Oxford Dictionary* as 'organized effort to achieve a particular purpose'. According to our scheme, drive is a necessary aspect of associative learning, of the remembered future. It is related to 'wanting', which we see as experiencing during the process of exploration towards attractor-related stimuli (e.g. food, security, water, sex). In terms of figure 4, wanting is the experiencing occurring during the process that leads from the gray CCS to the dotted one. Wanting implies a goal, even when the goal is not reached. There is persistence in the trajectory of change here, in the transition from the repulsor to the attractor state, and the sensory correlates of this transition are what we recognize as 'wanting'. The evolution of CSSs, an inevitable corollary of the evolution of associative learning in neural animals, gave rise to the individual, ontogenetic, learning-dependent animal telos – the will.

6. Discussion: the messy (evolutionary) biological approach to experiencing
'The really hard problem of consciousness is the problem of *experience*.' (Chalmers, 1995, p. 201, his italics).
'Consciousness is only a metaphysical term for the phenomena which are determined by associative memory'. (Loeb, 1900, p. 12).

Loeb's approach, which focuses on associative behavior and memory and identifies them with consciousness, seems very remote from that of Chalmers, who highlights the subjective, experiential aspects of consciousness. Chalmers accepts Levine's (1983) claim that there is an explanatory gap between functions and experience, and calls the bridging of this gap the "hard problem" of consciousness. He suggests that consciousness needs to be explained in terms of a physical primitive, like mass or electric charge, rendering a theory of consciousness more similar to elegant physics than to "messy biology" (Chalmers 1995, 1996). Other scientists (e.g. Hameroff 1994, 2007, Cairns-Smith 1996) suggest that physical principles beyond those used in today's biology (i.e. quantal processes), have to be incorporated and given center stage to solve the hard problem. We suggest that Loeb's focus on associative learning is essentially correct, although, as Chalmers emphasizes, satisfactory neurophysiological characterization of consciousness must, indeed, account for its intentional and subjective properties.

Understanding the biological foundations of consciousness in a way that will dissolve the mystery of these extraordinary neural processes requires multiple approaches. We are suggesting that useful insights may be generated by 'digging the tunnel' towards the biological understanding of consciousness from the evolutionary-transition end. We propose that the core property of associative learning in neural animals is the formation of memory-dependent, ontogenetically-constructed, integrated sensations and coordinated actions. Our focus on CSSs, which entail neural integration, is in line with other functional characterizations of consciousness (reviewed by Anil 2008). However, we believe that this integration, which is the foundation of both associative learning and experiencing, does not require highly evolved specific neural structures. It seems to occur mainly through classical Hebbian processes involving positive feedback, and through the re-entry processes highlighted by Edelman. We cannot rule out that additional mechanisms, such as transmission of signals among neurons through gap junctions, may also contribute to the integration that is essential for experiencing (Hameroff 2007).

Our proposal that motivating experiencing emerged within the context of the evolution of flexible associative learning has three major implications. The first and

most obvious one is that flexible associative animal-learning can be used as a behavioral criterion for an animal's experiencing. The second implication is that the neural processes underlying the simplest type of consciousness can be identified by comparing the neural organization in the simplest invertebrate animals able to flexibly learn by association with that of animals without such learning ability, or with only a very limited ability. For example, comparative studies of the memory mechanisms of cnidarians and ctenophores, which lack a CNS, with cephalized invertebrates, in conjunction with behavioral studies testing the learning abilities of individuals in these taxa, might shed light on the mechanisms involved. Our hypothesis predicts that experimentally induced selective inhibition of the proteins involved in associative learning will lead to selective loss of consciousness (but not to the loss of reflexes). The third implication of our proposal is that basic consciousness arose very early in animal evolution, when animals derived from small acoelomate flatworms (Baguñà 2008) acquired memory mechanisms enabling them to learn flexibly by association. This, we suggest, happened at the end of the Ediacaran or at beginning of the Cambrian, and was one of the factors driving the great adaptive radiation of the Cambrian (Ginsburg & Jablonka, 2007b; in preparation). The last conjecture departs from the view of most researchers, who assume that consciousness first appeared in the vertebrate lineage where affective responses are similar to ours; they believe that it evolved when the line leading to birds and mammals split from the ancestral reptilian line (Edelman & Tononi, 2000; Edelman, 2003), and that the thalamocortical system is essential for basic consciousness. Denton (2006), who suggested earlier emergence, and Merker (2007), who argued that the cortex is not necessary for the generation of primary states of consciousness, still focus on vertebrates.

There are not many modern studies that approach consciousness from the evolutionary end, and even fewer that attempt to account for its origins. Denton, one of the few biologists who has addressed the issue from an evolutionary perspective, emphasizes the importance of interoceptors, which signal to the animal whether or not it has departed from homeostasis (Denton 2006). Denton acknowledges the importance of exteroceptors that signal the state of the external world, but stresses the primacy of internal states. In our terms, the departures from homeostasis that Denton discusses are still instances of limited experiencing, for they are as yet functionless. We believe that for full experiencing to occur, both exteroceptors and interoceptors

must be involved – there is no point in being aware of one's internal states if one cannot do something about it. Humphrey (1992, 2000, 2006), who has approached the evolutionary origin of consciousness from a different angle, suggested on the basis of phenomena like blindsight that perception and sensation are distinct processes comprising consciousness. However, we believe that blindsight-like dissociations are already highly derived features of an evolved nervous system. We think that the evolution of what-happens-within-one (which Humphrey calls sensation) was achieved through learning what-happens-in-the-external-world (which he calls perception), and hence perception and sensation were mechanistically and evolutionarily linked from the outset. We share some of Dennett's (1991) evolutionary perspective, because just as he suggested in his multiple-draft theory, following repeated experiencing there are many potential, interconnected 'cryptic networks', one of which emerges from the ensemble of possibilities and becomes highlighted upon the right combination of stimuli.

Our view is closest to Edelman's position; he argues that the evolution of consciousness occurred within the context of relating complex integrated inputs to past learning responses and future needs (Edelman 2003). We believe, however, that the evolution of consciousness occurred much earlier than he suggested, and that it required less elaborate neural structures than the thalamocortical structures of highly evolved vertebrates. Our functional characterization also differs from that of Edelman, who stressed the importance of discrimination and perceptual image formation. Although we recognize the importance of discrimination and perceptual integration, we focus on motivation, which entails the integration and binding of remembered attractor-related inputs that form the remembered future.

There is no doubt that a transition to experiencing occurred during evolution, and hence that it has biological correlates that can be uncovered through comparative evolutionary studies. We have suggested that consciousness is a facet of the evolution of associative learning, and hence that understanding the evolution of the latter can yield insights for understanding the former. Animals remember and recall both repulsor-associated and attractor-associated cues, and we believe that the integrated dynamic sensory processes that accompany this learning and recall (the CSSs) are the simplest "unit processes" of consciousness, because they have properties compatible

with those attributed to primary conscious states. Learning through experiencing endows animals with motivation, so the evolution of experiencing altered animal evolution. It became dominated and guided by learning and by future goals.

Acknowledgments

REFERENCES

Anil, S. (2008), 'Functions of consciousness', in W. Banks (Ed.), *Elsevier Encyclopedia of Consciousness* (Oxford: Elsevier), in press.

Baguňà, J & Riutort, M. (2004), 'The dawn of bilaterian animals: the case of acoelomorph flatworms', *BioEssays* **26**, pp. 1046–1057.

Baguňà, J., Martinez, P., Paps, J., & Riutort, M. (2008), 'Back in time: a new systematic proposal for the Bilateria', *Philosophical Transactions of the Royal Society, B*, **363,** pp. 1481–1491.

Bennett, M.R. & Hacker, P.M.S. (2003), *Philosophical Foundations of Neuroscience* (Oxford: Blackwell Publishing).

Budd, G.E. (2008), 'The earliest fossil record of animals and its significance', *Philosophical Transactions of the Royal Society, B*, **363,** pp. 1425–1434.

Cairns-Smith, A.G. (1996), *Evolving the Mind: On the Nature of Matter and the Origin of Consciousness* (New York: Cambridge University Press).

Catling, D.C., Glein, C.R., Zahnle K.K., & McKay C.P. (2005), 'Why O_2 is required by complex life on habitable planets and the concept of planetary "oxygenation time" ', *Astrobiology*, **5,** pp. 415–438.

Chalmers, D.J. (1995), 'Facing up to the problem of consciousness', *Journal of Consciousness Studies*, **2**(3), pp. 200–219.

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

Chalmers, D.J. (2004), 'How can we construct a science of consciousness?', in M. Gazzaniga (Ed.), *The Cognitive Neurosciences III* (Cambridge, MA: MIT Press).

Damasio, A. R. (1994), *Descartes' error: Emotion, Reason and the Human Brain*, (NY: Grosset/Putnam).

Dennett, D.C. (1991), *Consciousness Explained* (Boston: Little, Brown).

Denton, D. (2006), *The Primordial Emotions: The Dawning of Consciousness* (Oxford: Oxford University Press).

Edelman, G. (1989), *The Remembered Present: A Biological Theory of Consciousness* (New York: Basic Books).

Edelman G. (2003), 'Naturalizing consciousness: A theoretical framework', *Proceedings of the National Academy of Sciences of the USA*, **100**(9), pp. 5520–5524.

Edelman G.M. & Tononi, G. (2000), *The Universe of Consciousness: How Matter Becomes Imagination* (New York: Basic Books).

Emes R.D., Pocklington, A.J., Anderson, C.N.G., Bayes, A., Collins, M.O., Vickers, C.A., Croning, M.D.R., Malik, B.R., Choudhary, J.S., Armstrong, J.D. & Grant, S.G.N. (2008), 'Evolutionary expansion and anatomical specialization of synapse proteome complexity', *Nature Neuroscience*, **11**, pp.799–806.

Fike, D.A., Grotzinger, J.P., Pratt, L.M. & Summons, R.E. (2006), 'Oxidation of the Ediacaran ocean', *Nature*, **444**, pp. 744–747.

Ginsburg, S. & Jablonka, E. (2007a), 'The transition to experiencing: I. The evolution of limited learning', *Biological Theory*, **2**(3), pp. 218–230.

Ginsburg, S. & Jablonka, E. (2007b), 'The transition to experiencing: II. The evolution of associative learning based on feelings', *Biological Theory*, **2**(3), pp. 231–243.

Hameroff, S.R. (1994), 'Quantum coherence in microtubules: A neural basis for emergent consciousness?', *Journal of Consciousness Studies*, **1**(1), pp. 91–118.

Hameroff, S.R. (2007), 'The brain is both neurocomputer and quantum computer', *Cognitive Science*, **31**, pp. 1035–1045.

Humphrey, N. (1992), *A History of the Mind* (London: Vintage).

Humphrey, N. (2000), 'How to solve the mind-body problem', *Journal of Consciousness Studies*, **7**(4), pp. 5–20.

Humphrey, N. (2006), *Seeing Red: A Study in Consciousness* (Cambridge, MA: Harvard University Press).

James, W. (1890, 1950), *The Principles of Psychology* (New York: Dover).

Levine, J. (1983), 'Materialism and qualia: The explanatory gap', *Pacific Philosophical Quarterly*, **64**, pp. 354–361.

Loeb, J. (1900), *Comparative Physiology of the Brain and Comparative Psychology* (New York: G.P. Putnam's Sons).

Mackie, G.O. & Meech, R.W (2000), 'Central circuitry in the jellyfish *Aglantha digitale*. III. The rootlet and pacemaker systems', *Journal of Experimental Biology*, **203**, pp. 1797–1807.

Merker, B. (2007), 'Consciousness without a cerebral cortex: a challenge for neuroscience and medicine', *Behavioral and Brain Sciences*, **30**(1), pp 63–81.

Nielsen, C. (2008), 'Six major steps in animal evolution: are we derived sponge larvae?', *Evolution and Development*, **10**(2), pp. 241–257.

Razran, G. (1971), *Mind in Evolution: An East-West Synthesis of Learned behavior and Cognition* (Boston: Houghton Mifflin).

Searle, J.R. (2004), *Mind – A Brief Introduction* (Oxford: Oxford University Press).

Smith, H.W. (1959), 'The biology of consciousness', in C.McC. Brooks & P.F. Cranfield (Eds.), *The Historical Development of Physiological Thought*, pp. 109–136 (New York: Hafner Publishing).

Spencer, H. (1855), *The Principles of Psychology* (London: Longmans).