Consciousness as a Biological Phenomenon: An Alternative to Panpsychism [1]

Catherine Wilson
University of York (UK)
The Graduate Center, CUNY

Forthcoming in *Harvard Review of Philosophy*.

Abstract: Following centuries of Cartesianism, which ascribed conscious awareness to humans alone, philosophers have begun to explore the possibility that experience in some form is widely distributed in the universe. It has been proposed that consciousness may pertain to machines, rocks, elementary particles, and perhaps the universe itself. This paper shows why philosophers have good reason to suppose that experiences are widely distributed in living nature, including worms and insects, but why panpsychism extending to non-living nature is an implausible doctrine.

The view that nonhuman animals lack experiences was current in the middle of the last century, but the scientific consensus has since altered dramatically, returning to much earlier assumptions regarding the extent of consciousness in nature. That mammals and birds and perhaps even cephalopods and insects are aware--in a sense yet to be clarified-- of an external world is increasingly accepted by ethologists.[1]

Some philosophers have pushed this trend further to the extent of arguing for some form of phenomenological awareness in non-living entities, including elementary particles and perhaps even the entire universe itself. In 1996, David Chalmers maintained that a thermostat might be conscious and that consciousness might even be associated in some way with a rock. "Some may worry," he commented, "about the fact that a thermostat is not *alive*, but it is hard to see why that should make a principled difference."[2] More recently, Chalmers has voiced a preference for idealism over panpsychism,[3] but panpsychism is alive and well in

[1] Acknowledgements: I am grateful to the audience of the History and Philosophy of Science Department in Cambridge where this paper was first presented in 2015, and especially to Hasok Chang for helpful criticism. I am also grateful to Lenny Moss for discussion of this topic, and to other participants in the New Materialism project at Rice University in 2013-4 headed by Sarah Ellenzweig and John Zammito for their ideas and stimulating discussion. Lars Chittka and Tim Fieblinger have also provided valuable input.

2018, with active debate amongst adherents over whether tables, chairs, rocks, electrons might be conscious.[4 5]

I argue here that consciousness is a feature of certain mobile, self-propelled, self-protective biological individuals—that is to say, a feature of *animals*. There is a principled difference between animals, which are products of natural selection, and other kinds of entities. Although synthetic animals with consciousness may someday exist, we probably don't have any yet, and neither a thermostat nor an electron will qualify as one. Meanwhile, the more we learn about the competencies of animals belonging to other taxa, including social insects, the more compelling it becomes to restrict consciousness to living entities.

In what follows, I will first sketch the metaphysical and epistemological background to the emergence of the "new" panpsychism, next discuss the results of experimental and ethological approaches to understanding consciousness, and finally explore and defend the position that we should understand consciousness as an adaptation for addressing the problems of living. This position blocks the inference to the very possibility of consciousness in most (though not all) non-living entities.

I

In the next to last chapter of *The Conscious Mind,* which I will take as a founding text of the new panpsychism, along with Galen Strawson's 'Realistic Monism,' Chalmers considered the question where, on the scale from fish, whose phenomenal awareness of a world he thinks is not in doubt, to slugs, to neural networks, to thermostats "consciousness should wink out." Consciousness for thermostats is, he suggested there, "one reasonable way for things to go, and on reflection perhaps as natural a way as any."[6] The thermostat, he said, will not be intelligent, it will not think, it will not have psychological properties like learning and wanting, but what he called "phenomenal consciousness" is something else, and the thermostat may have that.

Having phenomenal consciousness--sometimes referred to as having *qualia*--means simply having "experiences." We humans perceive identifiable objects such as tables, chairs, and cups of coffee, with their shapes and colours, scents and sounds, their movements and felt surfaces and textures. We see and sometimes recognize other people and animals. We experience "inner" sensations such as pain, itches, hunger, and fear, and we observe events involving people and things unfolding in a

world "outside" our bodies. If other entities have phenomenal consciousness, they must have something "like' what we have, though it is of course impossible to know exactly "what it is like" to be them.[7] "Someone who finds it crazy," Chalmers stated, "to suppose that a thermostat might have experiences at least owes us an account, of just *why* it is crazy." [8] Indeed, he went on to say, if there is experience in thermostats,

> there is probably experience *everywhere,* wherever there is a causal interaction there is information and wherever there is information (in for example the expanding and contracting states of a rock or in different states of an electron) there may be experience, not necessarily experiences *had* by the rock or electron but somehow associated with it.[9]

Panpsychism, or a closely related version of consciousness-attribution, is taken to be a consequence of the "hard problem,"[10] which is that there can be no intuitively satisfying answer to the question: How are experiences—perceptions, emotions, feelings—manufactured as the end products of a processes involving at their origins or somewhere along the way, neurological events?

The hard problem looks insoluble because by "explanation" we typically understand the recitation of a causal sequence that can be visualized to fill in the gap between two observable events. But we cannot even imagine a causal sequence beginning with a neural event and ending with an experience, even if common sense posits a definite cause-effect relation. We could discover that when, and only when their 'C-fibres' fire, subjects report being in pain. Or that whenever a particular neuron in her cortex is stimulated, a subject reports thinking of her grandmother. But no instrument, and certainly not our own eyes, can ever show us the fibres or the neuron producing the experience. [11] Producing experiences is not like producing cornflakes in a cereal factory, starting with raw ingredients.

To be sure, new types of entity can emerge from bases that are completely unlike them. Snowflakes, for example, are very surprising products of ordinary water vapour subjected to cold. Life arises from mixtures of carbon, oxygen, nitrogen, and other chemical elements. Galen Strawson argues that experience cannot be like snowflakes or life, an emergent phenomenon wholly dependent on an underlying non-experiential physical substrate. In his view, "For any feature Y of anything that is correctly considered to be emergent from X, there must be something about X and X alone in virtue of which Y emerges, and which is sufficient for Y." The way Y

emerges must, he says, be intelligible rather than brute and so miraculous. [12]

Strawson probably means by this slightly obscure formulation that we *can* understand how the cold forces the crystallization of molecules of water vapour into six-pointed shapes, and how, for example, from the behaviour of individual birds, the behaviour of a flock, which moves as a unit emerges. In these cases, physico-chemical things emerge from other physico-chemical things. But we can't in principle understand how the physico-chemical cells of our brains could act in concert so as to enable a (non physico-chemical) experience to emerge. Strawson concludes that the substrate for human consciousness must consist of conscious microphysical entities with their own form of microawareness, so that like arises from like.[13] Unlike Chalmers, however, Strawson rules out consciousness in macroscopic inanimate objects of our everyday experience.[14]

Chalmers maintained that researchers in the neurosciences are trying to solve only the "easy problems" of consciousness: "How does the brain process environmental stimulation? How does it integrate information? How do we produce reports on internal states?" They cannot answer the question, "Why is all this processing accompanied by an experienced inner life?" [15] And this explanation gap is the driver behind Chalmers's willingness, as well as Strawson's willingness, to extend consciousness to inanimate objects of our everyday experience.

The 'hard problem' is thereby connected to the 'zombie' hypothesis that surfaces repeatedly in *The Conscious Mind*. The zombie hypothesis is the proposal that a person could have a zombie-twin behaviorally indistinguishable from him or herself who is able to evince exactly the same competences in learning, reminiscence, and linguistic and social interaction, whilst lacking phenomenal consciousness. The thought experiment is supposed to show that having a certain physiological make-up, including a physiologically active human-type brain, is not sufficient for consciousness. But if consciousness is something extra that would have to be 'added' to a zombie to get a human, why not suppose that it could be 'added' to other entities besides zombies? Thus, the claim that a human brain is not *sufficient* for human-type consciousness gets converted into a claim that a brain of any sort is not even *necessary* for some form of consciousness.

To the empiricist, panpsychist conclusions about the possible distribution of experience and the reasoning that delivers them will seem strange. Phenomenal consciousness exists on the earth and perhaps on some other planets, or somewhere

else in the universe. But how could what we can report imagining happening in some possible world, namely, the existence of a twin zombie, inform us about the possible extent of consciousness in this universe, the one that we live in? We can imagine a world in which pebbles, or protons, or thermostats are conscious, as we can imagine a world in which nonhuman animals converse with one another at length in Flemish. But just as we should not infer from our imaginative abilities that maybe some of the actual nonhuman animals on Earth converse with one another at some length in Flemish, we should not infer that protons and thermostats might be conscious because we can imagine a world in which they are.[16]

II.

Both zombie arguments and the explanatory gap argument invoke the Cartesian idea of a simulacrum of an organism, but use it to argue for an un-Cartesian conclusion. This strategy has its perils.

Before Descartes proposed his beast-machine hypothesis, philosophers including Erasmus and Montaigne, as well as nonphilosophers, supposed that animals had feelings, contemplated options, made decisions, and experienced the world.[17] Descartes claimed to the contrary that animals were nonconscious mechanical devices, and he implied that a perfect simulacrum of a human being could be constructed that would only give itself away if it were put into a position of having to carry on a conversation or perform some intellectually challenging feat. Awareness, linguistic competence, reasoning, volition, and behavioral innovation or flexibility were human competencies he could not explain in "mechanical" terms, but since nonhuman animals did not appear to him to have these competencies, this was not a problem for his theory. He accordingly ascribed consciousness to entities that possessed them and declared that these competencies could not be explained mechanically but only by the possession of a separate, incorporeal soul. [18]

Today, zombie arguments can appeal to technological progress in AI. Electronic devices are becoming better and better at simulating linguistically competent, reasoning, inventive human beings, who can carry on a sensible conversation, write a poem or a musical composition, win a chess game, or drive a car. The power of unconscious mechanisms to produce lifelike behavior was explored theoretically by Valentino Braitenberg, who presented designs for what he termed "vehicles," beginning with a simple toy consisting of sensors wired to wheels that

could be made to approach or avoid or hover near a stimulus. [19] Braitenberg showed how not only context-appropriate actions, but what we think of as internal states such as memory and expectation, and even character traits such as optimism and egoism could be simulated, if what he called "Ergotrixwire," permitting learning of causal relationships and "Mnemotrixwire," for the laying down of memories, could be used in the construction of the vehicles. Although neither type of wire exists, some capabilities that were resistant to simulation when Braitenberg published his book in 1986, namely automated facial recognition, and linguistic dialogue, have substantially advanced, as has automated theorem proving in mathematics.

The original Cartesian argument depended on the principle of parsimony. If animal behavior can be explained without ascribing consciousness, the argument said, don't ascribe it, because the notion of the animal mind is a superfluous multiplication of entities. The Cartesian argument is:

> The technician can *roughly* explain/give a mechanical model for/simulate animal behavior without appealing to consciousness.
> In theory, animal behavior can be *perfectly* explained/simulated without appealing to consciousness.
> So, ascribing consciousness to animals to explain their behavior would be gratuitous, and should be avoided.

This argument looks pretty good. But…we don't argue as follows, even on the basis of the astonishing progress of simulating devices which may some day be able to do everything that humans do, including using language responsively and creatively, which Descartes took to be categorical evidence of a special human mind not possessed by other animals.

> The technician can roughly explain/give a mechanical model for/simulate my behavior without appealing to consciousness.
> In theory, my behavior can be perfectly explained/simulated without appealing to consciousness.
> So, ascribing consciousness to me to explain my behavior would be gratuitous, and should be avoided.

This argument looks pretty good too, except that the conclusion is unacceptable. I *know* that I am conscious, so parsimony gets no traction whatsoever. But if we refuse to be parsimonious in all contexts, consciousness gets detached conceptually from any sort of behavioral capability. Once it is detached, it can get attached to thermostats, electrons, and so on albeit in a very un-Cartesian manner.

To avoid being forced down this thought pathway, we can argue against Descartes's conclusion about nonhuman animals as follows:

> Even if I were nonlinguistic, never proved or invented anything, and was only as flexible in my behavior as the average cat or dog, and even if neuroscience or AI could explain or simulate all my behavior, I would still be conscious.
> Therefore, there is no reason to *deny* consciousness in the average cat or dog even if we can give perfect mechanical explanations or AI simulations of their behavior.

Indeed, denying consciousness to nonhuman animals is not really parsimonious. It adds a brand-new category of entities: animals *with* consciousness as opposed to animals without, whereas before we just had animals --and they were all assumed conscious until proven otherwise.

Further, we can go on to argue against Chalmers as follows: If I were an entity that not only was nonlinguistic and never proved or invented anything, but was only as flexible in my behavior as the average thermostat, it would be reasonable to conclude that I was unconscious. Denying consciousness to thermostats is as reasonable as denying consciousness to corpses. Their behavior, their range of capabilities, are *just too different* to that observed in known and imputed cases of consciousness. And that is a sensible, though still incomplete, answer to Chalmers's question why it seems crazy to ascribe consciousness to a thermostat.

III.

The observation that we have good reason not to take the possibility of consciousness in thermostats seriously nevertheless leaves the "hard problem" untouched. In addition to the designated "easy problems"-- information processing, or the identification of the neural correlates of experience-- what can the neurosciences actually tell us about phenomenality? In 1989, Stewart Sutherland, in a passage quoted by Chalmers in his Introduction, stated that "Nothing worth reading has been

written about [consciousness]." [20] That may have been true in 1989, but it could hardly be said today, and in this section I recapitulate some of the most significant finding and speculations about phenomenality stemming from the experimental sciences.

Three areas in particular, not coincidentally those Descartes himself identified with the possession of an incorporeal "soul," have been the focus of experiment and reflection. They are: the notions of a "self," "qualia" and "volition." [21] Light is shed on these notions largely through the study of aberrations--mistakes of consciousness --in people who are impaired though alive and semi-functional. They suggest answers to the question, "What good is consciousness to animals?" If we can understand why consciousness is good for animals, we can infer, with some confidence, that it was produced and enhanced by natural selection, and, further, that most entities that were not produced by natural selection probably lack consciousness. We cannot be sure that it was so produced. For example, having weight is good for animals since they do not fly off the surface of the earth, but everything that has mass has weight; animal weight was not produced by natural selection as understood by biologists. Nevertheless, animals of a given species have evolved to have a particular range of body weights--not too large, not too small--that is fitted to their way of life, and we can well ask how an elephant or a mouse uses its size to advantage.

We speak quite naturally of animals as doing things *for* or *by* or *to* themselves. They seek food for themselves, groom themselves, defend themselves, and mutilate themselves in captivity. A tipped over turtle struggles to turn itself right-side up; flies clean their legs, and crustaceans favour their wounded limbs. Animals also have extended selves; they relate possessively and in a caretaking way to things that are theirs, such as territories, mates, nests, and offspring.

All this implies a need to know where different parts of the body are, what they are doing, what hurts or feels nice, with respect especially to the significance of the information for survival and reproduction. Knowledge of mine, not mine, and someone else's with regard to external things is delivered in the feelings experienced in their presence or absence, not simply by perceptual recognition when they are in front of us. The features of organisms that supply what Antonio Damasio calls a "blueprint and anchor for a [conscious] mind" [22] appear to be: a boundary; an internal structure; a dispositional arrangement for the regulation of internal states; a narrow range of variability of internal states." As Richard Dawkins points out, in addition to

being "reactive, as a unit to changes within and outside its boundary," which could be said of a rosebush, a sentient organism must also possess one nervous system and move as a whole. [23] From a biological perspective, there is no point to experiencing pain and pleasure or having sensory experience if movement is impossible or if the animal is at the mercy of the wind or ocean waves and currents. Passive sensory consciousness is imaginable in the brain-in-a-vat scenario and actually occurs in people with locked in syndrome who can't move, but it can't be found in nonpathological nature. Corals, colonial animals fixed to the ocean floor, are unlikely to experience pain insofar as they are rooted to the spot and cannot avoid noxious stimuli.

Pathologies of disavowal bring out, by contrast with the normal case, what it is like to be a self, and why better, as opposed to worse awareness of one's own body and its relationships to other things can be critical for survival. A stroke patient may ignore everything on the left side of her body, with obvious potentially harmful or even fatal consequences: ignoring food on one side of her plate, or failing to notice approaching automobiles. Other brain damaged patients become convinced that their own arm or leg belongs to someone else and try to throw it out of the hospital bed or demand its amputation. Patients who believe themselves to be dead—Cotard's syndrome—do not relate to their bodies in the same fond and caring way as those who feel themselves to be alive.

Noting that children born without limbs seem to be prewired for experience of the missing limb, and that feelings of ownership and the illusion of control as well as pain persist in the phantom limbs of amputees, Ronald Melzack[24] asserts that the brain of a human being (and by implication that of many other animals) already represents a body typical of the species at birth.

Yet this bodily sense is also plastic: The sense of "my arm" fills a prosthesis once fitted. Ramachandran showed how to generate the illusion that one's own nose is located a few feet away, and by generating the illusion with mirrors that a patient could move his paralyzed arm with a painfully clenched fist, he was able permanently to relieve the discomfort. [25]

Other delusions of ownership and disavowal pertain to the "extended self." Feelings of "me" and "mine" are normally produced and maintained within a certain band that is neither too narrow and exclusionary nor too wide. People whose awareness of their own body is distorted and or who do not experience a sense of

ownership can die of anorexia, or set themselves on fire. In Capgras syndrome, close relatives are considered to be imposters because the feeling of their being "mine" is not aroused by their presence. Some demented elderly fail to recognize their spouses of many years, or, correspondingly, believe that unrelated strangers are people they know well. Schizophrenia involves intrusive thoughts that are experienced as control by minds or voices that belong to someone else—often God or the Devil.

Ownership of actions is also subject to distortion and manipulation. I can be made to feel that I am doing something I am not—moving another person's arm or controlling prices displayed on a computer by pressing buttons-- or that I am not doing something that I am. In table turning, participants in a séance are unaware that they are exerting force, and they ascribe the movement of the table to unseen spirits [26]

Note that the "sense of self" that is demonstrated indirectly by pathology and manipulation is not, as in classical metaphysics, either a pure intuition of "self to itself now" as Locke described it,[27] or referent of the Cartesian thought "I exist," or a function of autobiographical memory. Damasio describes the sense of self as "a second order mapping of the organism in the act of knowing, which is a feeling." [28] The feeling is not, however, a feeling of existence and continuity. Rather, consciousness is a feeling of knowing what is happening. "A conscious organism knows that it is relating to some object, and that the object to which it is relating is causing a change in it. "[29] We "construct and internally exhibit a specific kind of wordless knowledge—that our organism has been changed by an object."[30] A complex representation of the body thereby becomes itself an object of representation in its relations to an outside (spatially separated) or a temporally separated reality.

When consciousness is intact, I can distinguish my body from other bodies, my home and my offspring from other people's, and my actions from things just happening to me. The "sense of self" implies a corresponding sense of "not-self," and so the existence for me of an external world. [31] The external world of the dog with its sensitivity to odors, of the bird of prey with visual acuity and sensitivity to movement, are qualitatively different from ours, as are the worlds of the mole or worm living in the dark. The qualitative or phenomenological worlds of musicians and perfumers, whose sensory systems may have started off with advantages and which have been developed by long training are different from mine. [32] Since our visual system are each a little different, we can assume that we all see colors somewhat differently. Nevertheless, experience presents, irresistibly to us (what we call) "the world" outside

our bodies, and a body that is our own, but whose parts that we can do things to (lift, cut, see) as we do things to other external objects. How widely distributed in living nature might this capacity be?

Here Chalmers –and Spinoza--can be supported against Descartes. Take the earthworm—a bounded semipermeable soft bodied thing that is trying to survive long enough to produce more little worms. Its bodily state must be regulated so that it is not too hot, too cold, too dry, or too moist; it must try to avoid drowning, freezing and baking. While it cannot regulate its body temperature, it maintains internal homeostasis and tries to stay alive, though thanks to the existence of concrete pavements our modern worms are often unsuccessful in deploying their age-old repertoires. It has a central nervous system, with sensory cells, including some along its muscles that are probably responsible for proprioception.[33]

Are earthworms aware of their surroundings? They do not possess a cerebral cortex, the part of the human brain involved in memory, learning, thought, language, and consciousness. However, other animals may use other anatomical devices to manufacture worlds. A recent article in *Cell* discusses the marine worm *Platynereis dumerilii*. [34] This worm, according to its investigators, "lives in self-made tubes, explores its environment actively for food, and shows signs of learning behavior." When it evolved, the authors explain in interview, "the seafloor would have contained various food sources. "In order for organisms to explore these foods, it would have been "advantageous to evolve a brain center that was able to integrate the different smells and ultimately learn what is good and what is bad food."

The "mushroom bodies" in invertebrates such as insects, spiders, crustaceans and velvet worms, according to the authors, "serve as a center for associative learning and memory formation, activities that are very similar to those of the cerebral cortex." They have been shown to be responsible for associative learning, and their discoverer Félix Dujardin proposed in 1850 that they gave these tiny creatures free will, an ability to act against instinct. [35] As Griffin argues, "[T]he capability of conscious awareness under some conditions may well be so essential that it is the sine qua none of animal life, even for the smallest and simplest animals that have any central nervous system at all. When the whole system is small, this core function may therefore be a larger fraction of the whole." [36] Chittka remarks of the bee which co-operates to build an elaborate geometrical structure, the hive, evaluates food sources, maps its territory, remembers threats, communicates information to its fellows,

provisions its young, enters into seemingly emotional states, and can solve novel problems under laboratory conditions, that references to 'instinct' fail to do justice to the behavioural complexity of social insects. [37]

But why have qualia? Gerald Edelman and colleagues argue that the organism needs to be faced with a scene in order to extract and employ information. "Even implicit learning initially requires consciousness of stimuli from which regularities are unconsciously developed."[38]

It is helpful to remember in this connection that the human brain is a system of 160 billion neurons and supporting cells in interaction, interacting at tremendous speeds, far faster than today's supercomputers. There is nothing stable in these fleeting, flicking configurations, regardless of whether we are asleep or awake: even under anaesthesia, our brains are running. Yet we see a world of objects, many of which are fixed in place, have determinate properties, and can move at trackable speeds. Ramachandran and Hirstein argue that qualia are stable and "irrevocable"[39]. The visual scene is a decision about how reality is that is a selection out of multiple possibilities. What makes decisions about what's in the world necessary is that the physical stimuli, light waves, for example, that reach our eyes (elements of our common, nominal world) are consistent with a number of different scenes. Even for the individual animal, there is no unique mapping from these physical signals to experience. As they elaborate the point, "To deliberately overstate the case it's as though when you look at even the simplest visual scene, you generate an endless number of hallucinations and pick the one hallucination that most accurately matches the current input."[40] When possibilities are equiprobable, and even when they are not, consciousness may select the most optimistic or pessimistic one, or select in accord with some other heuristic such as recency.

The world-making mechanisms of the brain and body, then, serve to present a scene, inviting action or permitting inaction. The nervous system conjures up a preferred presentation, visually, acoustically, in terms of taste, smell, or touch, or in terms of the other senses we lack, thereby deciding what entities there are in the environment, how far away they are, what is occluding what, how dangerous or salutary they are. Consciousness presents to the animal a world in which it can exercise, and fail to exercise, or exercise maladroitly or successfully, control, in situations calling for decisions, elegantly described by the Dawkins as events which lead to "a sudden decrease in the uncertainty of future behavior."[41] A decision might

indeed be regarded as the psychological analogue of homeostasis, according to which the organism seeks to keep its future experience within a narrow range of envisioned bounds. To do so, it may compare "pictorially" the present with the past and with an imagined future.

Information about what, where, how fast, how strong, how sturdy, mine or not mine, as well as the qualities of the whats—how ripe, how edible, how friendly, how needy—is delivered, not on a narrowly "need to know basis," but on a far broader one, perhaps making consciousness seem like an overly generous endowment. That is, I don't need right now to see that the tree outside my window has thousands of individual leaves, but either I do sometimes need that level of perceptual detail or my ancestors did and I have retained an unused capacity.

Pathologies involving qualia include not only blindness, deafness, lacking a sense of smell or taste, for which the disadvantages are obvious, but other partial defects. Stroke patients suffering from "blindsight," studied by Larry Weiskrantz,[42] though they can perform a range of tasks including navigation, (some can even drive a car), and though they can register the emotional significance of faces and depicted scenes, cannot do other things usually necessary for survival and reproduction, such as determining the age or gender of those faces. Faced with either a horizontal mailbox or a vertical mailbox the blindsighted person can put the letter in, but faced with a horizontal mailbox and a vertical mailbox they don't know what to do. These findings reinforce the view that consciousness is sometimes needed to make decisions. Its invention in evolutionary history would seem to parallel the invention of freedom in the sense of having, in your species-repertoire, different possible responses to sensed external and internal conditions. Consciousness is arguably a condition of behavioral flexibility, another early insight of Descartes.

At this point a pertinent objection arises to the claim that consciousness is necessary for at least some forms of life. Couldn't a well-programmed machine succeed in the mailbox task, the recognition of male and female faces, and so on? We know that activities such as orienting, crouching, flinching, pouncing, are efficiently automated.[43] A frog will snap at anything of a given size, shape and distance; the costs of wasting energy or capturing unwanted prey are outweighed by nutritional benefits. Some simple heuristic similar to that governing the frog's responses may even determine whether a lioness attacks a gazelle or conserves her strength for a

better opportunity. We find ourselves getting out of bed or out of the bath without having made a conscious decision that now is the time.

Since the original research on blindsight, revealing not only what patients cannot do, but what they can do, it has become apparent how much of human life is accomplished on a subconscious level. Motor activities such as riding a bicycle and driving become automated or semi-automated. Tasks like hammering can be carried out without the subject's attention to what they are doing. We often write and speak "spontaneously" without having consciously decided what to say or write. The hand is removed from the flame before the sensation of pain is felt.

A person who is dysmorphic, blindsighted, who suffers from a neglect or disavowal syndrome, or a mixed-up sense of voluntary control, is clearly at a survival and reproduction disadvantage, though the deficiencies may be compensated for or made irrelevant in the context. We can indeed believe that a heritable tendency to these conditions would not be preserved and that the underpinnings for not having them –for having a normal self--are heritable. But the problems of these people, it might be argued, arise from defects in their nonconscious processing. Impaired consciousness, it might be argued, always indicates impaired information processing on a nonconscious level. Therefore, we gain no insight, according to this objection, into the advantages of consciousness from considering defects in conscious experience.

Indeed, contrary to what Ramachandran suggests, the automatic machinery might decide both what reality is and what to do about it. Perhaps some form of information processing decides that there is a tree in front of me, and I get a useless epiphenomenal presentation delivered to me in the form of the experience of seeing a tree. Some form of information processing completely unavailable to consciousness decides that I am going to climb it, and I get a feeling of volition that bears no causal relationship whatsoever to my subsequent climb.

My feeling that I decided to climb the tree and am doing it out of my own free will is –according to much recent experimental evidence—just the information delivered to me that I am going to do, or doing this thing that my nervous system after summing and ranking and otherwise manipulating a slew of inputs is now set to do, with a feeling of ownership or "authorship" attached. *Seeing that* there is milk in the refrigerator together with my desire for milky coffee gives me a reason to—voluntarily—reach for it and pour it into my coffee. But all I can state with

confidence is that I saw the milk, that I recognized it as something I needed or wanted, that I reached for it, and that the action as it was happening and in retrospect felt like something I did because I wanted to do it and no one else was making me do it.

On this view, the feeling that we are making a decision or performing a voluntary action is the perception of a thought as the cause of an action, but in fact the action or decision is initiated –via what D.M. Wegner calls "unconscious and inscrutable mechanisms,"[44] before we are aware of the decision or volition. Once the unconscious processing mechanism has brought about an action, volitional feelings and later reasons appear in our minds for why we decided to do that. We come to believe that these reasons caused our choice. The will, says Wegner, is an illusion because consciousness has no input into actions. Wegner thinks that we need the illusion of free will because it is important for blame and the sanctions that maintain responsibility in social life.[45]

Only the neural correlates of conscious experience, it seems, can fully fill in the causal chain. We seem to be pushed and pulled by hunger, thirst, lust, aggression, and so on, but the seeings and realizings of phenomenal consciousness cannot initiate physical movements of our limbs, changing the quantity and direction of movement in the universe. One might object that if ordinary examples of deliberate activity do not establish that consciousness has causal powers, other examples do, including the voluntary control over some autonomic processes that Yogis can achieve, or the control over the firing of single neurons, that can be achieved with practice and biofeedback. [46] And don't mental states of preparedness make a difference to how we react, physically, to loud noises or bad news, or temptation? None of these phenomena, however, can reveal consciousness as causally active; they are consistent with epiphenomenalism.

Against this line of argument, one might ask why, if conscious awareness makes no difference to survival and reproduction, the nervous system goes to so much trouble to differentiate between things *I* did and things that happened to my body, inducing in me the "illusion" that I control my actions? Why couldn't everything be presented to me as a happening to my body? But it is vitally important for a squirrel or a mouse to know that *it* initiated some action, not for moral or social reasons, but because information about what outcomes it can control and what happenings it can't is important for survival. [47] And if our decisions and behavior were governed by

unconscious mechanisms, and if consciousness contributed nothing to our functioning, then, as Brian Earl argues, we could go through life dreaming or hallucinating and it would make absolutely no difference. If qualia were useless, he points out, it would seem unnecessary for the nervous system take such care in composing and adjusting –via mechanisms for color and shape constancy for example--the qualitative details in the "scene."

Further, the following activities, cited by Earl, are difficult to automate: "Interacting with conspecifics; attending to a sudden or unexpected event; paying close attention to a task; being alert in unusual, interesting, or unpredictable situations; observing events as they occur; imitating an action that is contrary to habit;" or learning a new skill. Even more difficult to automate are such activities as "mentally processing instructions, thinking through an expected event, or preparing actions by mental rehearsal, and remembering and reviewing events."[48] Yet these activities can be essential for survival in a range of birds and mammals, as well as in humans, whether or not they are needed in the other taxa.

IV.

The hypothesis I propose is that consciousness evolved in animals because certain activities that give them a survival and reproduction advantage are difficult to automate. To be sure, we can imagine these activities being fully automated in a possible world (the zombie scenario), and it is even possible that they could be automated in our world (the AI scenario). A very large robot, with super small chips or working on quantum principles might be able to do exactly what animals do. The robot would not only convert energy from the environment into purposeful activity; it would be able to make good, but slightly imperfect copies of itself, exhibiting enough variation for natural selection to work on in a demanding and changing environment. The hypothesis is simply that consciousness can produce the competencies in question more efficiently than an unconscious mechanism could. The unconscious device would require too much energy of a type difficult to obtainable in oceans, forests, and plains; would need too much complicated, buggy software, and too much massive, prone to breakdown hardware to carry out the activities of life.

The existence of consciousness in our world is, then, explained by the fact that it is useful to the animals that have it, just as their hearts, wings, habits of burrowing,

or habits of reciprocal altruism are useful to the animals that have them. We should be astonished if consciousness arose in non-animal entities that cannot use it because there would be no explanation at all for its presence.

It might be objected that nothing said so far favors the conclusion that *consciousness* was selected for, rather than some underlying machinery that --however complex or even mystifying in the eyes of human engineers--serves as its substrate and so is responsible for all animal competencies. If it is only certain forms of neural anatomy and physiology that were selected for, then, it seems, we cannot characterize *consciousness* as such as useful, and the argument that we should be astonished if consciousness were uselessly present in non-living entities is unavailable. We have no way to determine whether:

a) Natural selection favored phenomenal consciousness, because organisms with small degrees of phenomenal consciousness and a subjective sense of agency were originally favored over their less conscious competitors, and organisms with more consciousness were favored up to a point over competitors with less.

Or:

b) Natural selection favored certain competencies and gradually developed the machinery to support them in various forms and degrees by favoring variants that possessed them. These competencies happen to be correlated with phenomenal consciousness and feelings of agency, but these were not as such targets of selection.

But the impression of "choice" between a) and b) is misleading. Nature does not "favor" either consciousness or machinery. Nature simply drives down the frequency of genes belonging to individuals who encounter unfavorable life conditions and who die without outbreeding their conspecifics. The result has been that animals with particular neurological substrata making experience possible appeared on the planet, including perhaps worms and insects, but most certainly birds and mammals. We can infer from the ecological niche they occupy that worms don't feel love or have visual dreams with wild plots, but the niches occupied by birds and mammals readily allow

for this. The evolving neural morphology and evolving competencies of the included taxa can nevertheless as far as we know be traced back to the same common ancestor.

To conclude, from an ethological standpoint, it makes no sense to ask: Why would *my zombie-twin* benefit from being conscious? That is like asking, why would *this pigeon*, if it didn't have wings but could do everything a pigeon with wings can do, benefit from having wings? Or why would *this worm,* if it could do everything a tunneling worm could do but without tunneling, benefit from being endowed with a tunneling habit? The question is not how pigeons without wings came to have wings, or how birds, insects, and mammals that needn't have been conscious became conscious. The question is how we got pigeons, tunnelling worms, and conscious animals. The answer is that the evolution of consciousness, like the evolution of wings and hearts, created new niches in ecological space for new types of living individual. Wings permitted the organisms that evolved them to take to the air to find new sources of food such as pollens and insects and to evade ground dwelling predators. Hearts permitted the organisms that evolved them to distribute oxygen and nutrients and collect waste products; they were preconditions of the development of brains. In inventing the underlying mechanisms that are necessary and sufficient for awareness of an external world, for awareness of a self distinct from other objects of interest, and for awareness of this self's agency in that world, nature made the forms of life of humans and many other animals possible.

The thermostat doesn't have any of the internal mechanisms –a cerebral cortex or mushroom bodies -- found in entities known or reasonably believed to be consciousness. Nor is it partaking in the struggle for existence or competition to pass on its genes to the next generation. It isn't a member of a lineage that developed consciousness as a valuable trait. Above all, it doesn't move or make decisions and doesn't need to have a world presented to it. Accordingly, the answer to the question why it is crazy to suppose the thermostat could be conscious is that it is stuck to the wall.

There is one final important objection to the claim that by taking consciousness as an evolutionary invention, we can decisively rule out its possession by non-living things. The objection is that evolutionary inventions can pre-exist or be reproduced in non-evolutionary contexts and formats. For example, luminescence is an evolutionary discovery of fireflies and some marine animals, but light is also produced naturally in the non-living stars and can be produced artificially by humans

in lightbulbs. Analogously, evolution may simply have chanced to hit on a phenomenon--consciousness-- pre-existing in non-living nature and technologically reproducible. Like light production, consciousness production may have been 'discovered' for the use of animals when the demands on them for behavioural responsiveness reached a certain level of complexity. without being restricted to them. So the question whether consciousness is a naturally occurring phenomenon that can appear in a range of systems is a real one.

As suggested above, there is no reason to rule out the possibility of synthetic conscious life in the form of devices utilising new miniaturising technologies that essentially copy the features of animal systems that are necessary and sufficient for consciousness, including sensory organs and a unified nervous system. That said, we have good reason to doubt that a device consisting, for example, of the population of China equipped with two way-radios and satellite connectivity could have experiences of a changing visual, auditory, olfactory, etc. scene with the twin hallmarks of differentiation and integration.[49] We have even better reason to doubt that chairs and electrons, which have no sense organs and cannot move in purposeful ways, could be conscious of a world of things separate from themselves. Panpsychists frequently refer in this connection to "very simple" forms of consciousness, but that term is a mere placeholder for an I-know-not-what. They are not entitled to indicate by the term "very simple" the kind of consciousness we might imagine a worm or a wasp possesses.

Accordingly, in order to have good grounds for rejecting panpsychism, it is not necessary to solve the hard problem. We need only agree that it does not follow from our inability to solve that problem that consciousness could be found anywhere or everywhere. We do not need to know exactly *how* unconscious physical entities, in the right number and arrangement, doing what they normally do, produce consciousness in order to consider its possession by electrons or thermostats as a hypothesis not worth advancing, let alone pursuing. If you accept it that consciousness implies the existence of *experiences*--which, after all, was the launch pad for the "hard problem" -- and if you accept it that experiences are unlikely to exist where there are neither *decisions* relevant to survival and reproduction to be made, nor were, formerly, such decisions to be made (as in the case of locked-in syndromes), nor deliberately manufactured simulations of experiencing and deciding, you will have a good defence against an ambitious but undertheorized panpsychism.

Further, as I hope to have shown here, progress on the so-called 'easy problems' of consciousness (which no researcher into animal consciousness would describe in that way) can begin shed light on the many dimensions of conscious awareness in perception, volition, and cognition.

---

[11] Charles Darwin suspected that beetles and planaria were aware of their surroundings and made decisions. § 8-9 in "Old & useless notes about the moral sense & some metaphysical points." (1838-1840) Transcribed and edited by Paul H. Barrett. (*Darwin Online*, http://darwin-online.org.uk/) On the behaviour and putative mentality of cephalopods such as octopi and squid, see Peter Godfrey-Smith, *Other Minds* (New York: Farrar, Strauss, and Giroux, 2016). Beginning with his *Animal Thinking* (Cambridge MA: Harvard University Press, 1985) and *Animal Minds: Beyond Cognition to Consciousness* (Chicago: University of Chicago Press, 1992), Donald Griffin aroused consternation for views that are now mainstream. His proposal that insects might be conscious was regarded as particularly outrageous, but see Lars Chittka, "Bee Cognition,"*Current Biology* 27:19, R1049-R1053 (2017) and M.J. Sheehan and E.A. Tibbetts, "Specialized Face Learning is Associated with Individual Recognition in Paper Wasps." *Science* 334 (2011), 1272–1275. The question whether the competencies described actually *imply* consciousness as opposed to mere cognition is discussed below.

[2] David Chalmers, *The Conscious Mind: In Search of a Fundamental Theory* (Oxford: Oxford University Press, 1996), 294. See also Philip Goff, *Consciousness and Fundamental Reality* (Oxford: Oxford University Press 2017) for an account for an account of a conscious universe.

[3] Chalmers, 'Idealism and the Mind-body Problem,' Forthcoming in W. Seager, ed., *The Routledge Handbook to Panpsychism*, Abingdon, Routledge, 2018.

[4] See for example, Olivia Goldsmith, https://uncommondescent.com/intelligent-design/at-quartz-materialists-are-converting-to-panpsychism. Philosophers

interviewed for *Quartz* magazine were explicit: "One interpretation of the theory holds that 'any system is conscious' says Chalmers. 'Rocks will be conscious, spoons will be conscious, the Earth will be conscious. Any kind of aggregation gives you consciousness.'"

[5] Galen Strawson, "Realistic Monism - Why Physicalism Entails Panpsychism." repr. in *Consciousness and its Place in Nature*, ed. Anthony Freeman (Charlottesville, VA, Imprint Academic, 2006), 3-31.

[6] Chalmers, *Conscious Mind*, 292.

[7] As argued by Thomas Nagel in 'What is it Like to Be a Bat," *The Philosophical Review* 83: 4 (1974), 435-450.

[8] Chalmers, *Conscious Mind*, 293.

[9] Ibid. 295

[10] Strawson reminds his contemporaries that the "hard problem" is not a recent invention or discovery but was familiar to every major philosopher of the 17th century. It was never lost sight of, despite a movement to declare it off limits to science in the early to mid 20th century. "Stoppard's Real Hard Problem," *TLS*, No 5839, Feb 27, 2015.

[11] Leibniz pointed out that an organ of thought as big as a mill whose insides could be visited and studied would not show the curious tourist *how* consciousness was produced as an end product. *Leibniz: Philosophical Essays*, trans. Roger Ariew and Daniel Garber (Indianapolis: Hackett, 1989), 215. The argument is meant to show that consciousness pertains only to "simple substance" but it fails as a *reductio ad absurdum* of the very idea of a thinking machine.

[12] Strawson, 'Realistic Monism,' 18-19.

[13] Strawson, "Realistic Monism," 27. An earlier version of this essay is criticized in C. Wilson "Commentary on Galen Strawson," in *Consciousness and its Place in Nature*, ed. Freeman, 177-83.

[14] Strawson, "Realistic Monism," 26.

[15] Chalmers, *Conscious Mind*, ii.

[16] It's possible that some Belgian parrots have extended conversations with one another in Flemish; but we would have no reason to think so unless we observed them conversing with one another in at least some natural human language. Yet thought

experiments are frequently supposed in philosophy to reveal reality to us. Hume appears to argue that any sequence of events we can imagine taking place could take place: "Is there any more intelligible proposition than to affirm, that all the trees will flourish in December and January, and decay in May and June? Now whatever is intelligible, and can be distinctly conceived, implies no contradiction, and can never be proved false by any demonstrative argument or abstract reasoning *à priori*." *An Enquiry Concerning Human Understanding,* 4.18, ed. Selby-Bigge, rev. P.H. Nidditch (Oxford: Clarendon, 1975) 35.

[17] As Montaigne asks, 'When I play with my cat, who knows if I am not more amusing to her than she is to me?' *Essays*, Book II, Ch. 12. For details of respect for animals across a range of periods and cultures, see Rod Preece, *Awe for the Tiger, Love for the Lamb* (Vancouver, University of British Columbia, 2002).

[18] "[T]hey have no intelligence at all, and ...it is nature which acts in them according to the disposition of the organs," Descartes, *Discourse on the Method* in *The Philosophical Writings of Descartes*. 2 vols., ed. and trans. John Cottingham, Robert Stoothoff, and Dugald Murdoch, (Cambridge: Cambridge University Press, 1985). I: 141

[19] Valentino Braitenberg, *Vehicles: Experiments in Synthetic Psychology*, (Cambridge: MIT Press, 1986).

[20] Chalmers, *Conscious Mind*, 1.

[21] V. S. Ramachandran and William Hirstein, "Three Laws of Qualia: What Neurology Tells Us about the Biological Functions of Consciousness, Qualia and the Self" *Journal of Consciousness Studies,* 4: 5-6 (1997), 429–58.

[22] Antonio Damasio, *The Feeling of What Happens: Body and Emotion in the Making of Consciousness* (New York: Harcourt Brace, 1999) 136.

[23] Richard Dawkins, *The Extended Phenotype* (Oxford: Oxford University Press, 1982), 248.

[24] Ronald Melzack, "Phantom Limbs, the Self and the Brain," *Canadian Psychology/Psychologie Canadienne* 30 (1989) 1-15.

[25] V.S. Ramachandran and D. Rogers-Ramachandran, "Synaesthesia in Phantom Limbs Induced with Mirrors," *Proceedings of the Royal Society: Biological Sciences* 263:1369 (1996) 377–386.

[26] D.M. Wegner, *The Illusion of Conscious Will* (Cambridge MA, MIT Press, 2002) 7-9.

[27] Locke, *An Essay Concerning Human Understanding*, ed. P.H. Nidditch (Oxford: Clarendon, 1979). II:xxvii.10.

[28] Damasio, *Feeling of What Happens*, 136.

[29] Ibid. 133.

[30] Ibid.168.

[31] I adopt, but adapt, Ted Honderich's notion that where consciousness is, there a world is. "The difference between me now and a chair in this room, it can be said, is that for me a world exists, and for the chair a world does not exist. Or rather, as I prefer to say, *my consciousness now consists in the existence of a world.*" *On Consciousness*, (Pittsburgh, PA: University of Pittsburgh Press, 2004), 130. For criticism, mainly on syntactical grounds, see Colin Mc Ginn's review in *The Philosophical Review,* 116 (2007), 474-7.

[32] Citations in M. Bendsafi et al, ""Olfactory and Gustatory Mental Imagery: Modulation by Sensory Experience and Comparison with Auditory Mental Imagery," in Simon Lacey and Rebecca Lawson, eds., *Multi-Sensory Imagery* (New York: Springer, 2013), 77-92.

[33] C. Ladd Prosser, "The Nervous System of the Earthworm," *The Quarterly Review of Biology* 1934 9:2, 181-200, see. esp. 189-90.

[34] R. Tomer, R., A. Denes, K. Tessmar-Raible, and D. Arendt, "Profiling by Image Registration Reveals Common Origin of Annelid Mushroom Bodies and Vertebrate Pallium," *Cell*, 142:5 (2010), 800-809.

[35] Félix Dujardin, *The Nervous Systems of Invertebrates: An Evolutionary and Comparative Approach,* ed., O. Breidbach and W. Kutsch (Basel: Birkhauser, 1995) 350.

[36] Griffin, *Animal Minds*, 3.

[37] Chittka, "'Bee Cognition," 1049.

[38] Anil K.Seth, Eugene Izhikevich, George N. Reeke, and Gerald M. Edelman. "Theories and measures of consciousness: An extended framework." *Proceedings of the National Academy of Sciences of the United States of America,* 103: 8 (2006), 10799-10804, quotation on 10799. The "scene" is both differentiated, in containing what Kant would call a manifold of qualities in phenomenal space (and time) and

integrated in being referred to a unified, subjective point of view--Kant's unity of apperception (*Critique of Pure Reason*, A 106-7). Cf. Ramachandran and Hirstein, "Three Laws of Qualia," 453. "The single most important principle underlying the mechanisms of perception and conscious experience: that they may have evolved exclusively for extracting statistical regularities from the natural world."

[39] Ibid. 437.

[40] Ibid. 442.

[41] Marian Dawkins and Richard Dawkins "Decisions and the Uncertainty of Behaviour," *Behaviour* 45 (1973) 83-103.

[42] Larry Weiskrantz, *A Case Study and its Implications* (Oxford: Oxford University Press, 1986).

[43] Brian Earl, "The Biological Function of Consciousness," *Frontiers in Psychology*, 5 (2014), 1-18.

[44] Wegner, *Illusion*, 98. See also, Benjamin Libet, "Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action," *Behavioural and Brain Sciences*, 8 (1985) 529-566 and Max Velmans, "Is Human Information Processing Conscious? *Behavioral and Brain Sciences,* 14:4 (1991) 651- 669.

[45] Wegner, *Illusion*, 328-341.

[46] Max Velmans "How Could Conscious Experiences Affect Brains?" *Journal of Consciousness Studies*, 9 (2002), 3–29.

[47] As Moss argues, "[N]ature explores greater levels of detachment [from determination by the immediate internal or external environment] and … at increasingly higher levels of detachment this increasingly amounts to moving toward the capacity for normative self-determination." L. Moss, "Detachment Theory: Agency, Nature, and the Normative Nihilism of New Materialism," in S. Ellensweig and J. Zammito, eds., *The New Politics of Materialism: History, Philosophy Science,* (Routledge: NY, 2017), 225-47, quotation on 235.

[48] Earl, "The Biological Function of Consciousness," 10-13

[49] Ned Block, "Troubles with Functionalism," in *Minnesota Studies in the Philosophy of Science*, Vol. 9, ed. C. Wade Savage (Minneapolis: University of Minnesota Press, 1978), 261-325.