THE PROBLEM OF CONSCIOUSNESS

© 2019 by Dan Bruiger email: dbruiger [at] telus [dot] net

From the time that humans began thinking about their own thinking, understanding the nature of consciousness has posed a special challenge. This is why philosophers in recent times refer to it as the *hard* problem of consciousness—to distinguish it from the problem of understanding how the brain controls behavior, which is by no means easy. Technology, such as brain scanning, has greatly advanced understanding of the nervous system as a control device. But the question of how the brain, or any material thing, can *have experience* is altogether another sort of question. I will begin by talking about how and why this is different from usual scientific questions, and then go on to consider what sort of approach it requires because of that difference.

But first let's clarify what is meant by 'consciousness', because mental terms are notoriously ambiguous. Here it will mean the actual experiences one normally has while awake—but also during dreams, while hallucinating, while daydreaming, or during acts of imagining, etc. Philosophers often refer to this as *phenomenal* experience, or phenomenal consciousness. To use a single term instead of a phrase, let us call it *phenomenality.* The reason for this formal term is that the commonly used word, *experience*, is also ambiguous. On the one hand, experience could indicate the sensation in your jaw when you have a toothache or the anxiety you feel during a job interview: your here-and-now experience. On the other hand, it could refer to memory and to learning, such as the traumatic experience you had with a particular dentist or the accumulated work experience on your résumé. This is not a trivial distinction, because it underlines what renders consciousness different from ordinary scientific topics. And that difference is: point of view.

The toothache or the anxiety is only a part of what transpires in the dentist's office or during the job interview. For one thing, they seem to take place in one's body, whereas the dentist or the prospective employer, and their offices, clearly exist in the world outside the body. *All* experience has in common, however, that it seems to take place from the point of view of an observing self. I see the dentist's drill descending into my anesthetized mouth and feel in my jaw the vibrations of the drilling. (Without the anesthetic I would feel pain as well.) I am witness to these experiences. However, if I try to describe them all in the same terms, strictly as a series of events in the world, I am strangely unable to account in that language for the sensations in my body, even though I may think of my body as a part of the world and distinct from my observing self. My pain and other bodily sensations obviously depend on my physical body, which is part of the world. Yet, they seem to be a different sort of occurrence than events in the world such as the dentist's movements or the drill's operation, which I can well imagine to take place whether or not I personally am there to witness them. If I happen to be a neuro-scientist, I might believe that a theory of how the brain works as a control system could explain the behavior of the dentist's muscles, the chemical action of the anesthetic on my nerves, and so forth. I could also hope to explain the physics behind the operation of the electric drill and the air currents in the room. All of these are things going on in the dentist's office. But could I explain in such terms the pain of the toothache or the purposes of the dentist in trying to alleviate my discomfort, to earn a living, etc.?

The distinction concerned is so commonplace that it is built into our language. This is the difference between a *first-person* and a *third-person* point of view. (The "first" person, of course, is oneself, and the third person is "he, she or it.") Only "I" can experience the pain in my body, because no other perceiving subject is hooked up directly to this body's nervous system. Another perceiving subject—the dentist, for instance —is connected to my body in quite a different way. By means of his own nervous system, the dentist can witness the behavior of my body (such as the swelling of my gums and the symptoms of discomfort I display) and can listen to my reports of being in pain. But he cannot feel my toothache. Of course, the dentist can *imagine* my experience, having probably experienced his own toothache at some time. But he is also perfectly free to imagine that I have no such experience at all, since I cannot prove that any such thing as the sensation of pain (or any other sensation) even exists! Humanly speaking, that would be an outrageous attitude, perhaps pathological. But it is exactly such an embarrassment that is at the root of the hard problem of consciousness. While skeptical dentists might be a problem for society, they are no more than tools in the standard kit of the philosopher thinking about consciousness.

To come to the point: consciousness cannot be understood scientifically because right from the start science has excluded the first-person from its approach. It is only interested in describing the world from a third-person point of view, which others may share and corroborate. This simply reflects and extends our ordinary way of perceiving the world, which is largely visual and from a distance. Science explains things in terms of cause and effect. That means a description of how one thing in the world "over there" impinges on another—perhaps in an endless sequence of toppling dominoes. Unconcerned with how the world impinges on the subject, or the subject on it, this is a description from a third-person perspective. Furthermore, the ideal of scientific description is mathematical: a quantitative description of events in some "space," if not the space-time continuum of physics then some conceptual space. This too is third-person description, of events that are supposed to take place objectively and remotely in the external world.

Conducted on that basis, science has proven very effective at manipulating the natural world. This natural world includes the body itself as a topic for scientific study and manipulation. The body is an *object* external to the perceiving subject, who seems to be at its mercy in some ways and its master in others. But this focus on external objects precludes considering the influence the perceiving subject has on perception. In scientific method, this subject is idealized as the scientific observer, which is supposed to exclude such influence. But what it actually excludes is anything idiosyncratic in the perception of a *given* observer. Scientific description is not supposed to be an anecdotal account of personal experience. The feelings, whims, or raw imagination of the scientist have no official place in conducting experiments, reporting their results, or in formulating theories. But this approach, as fruitful as it is, does not address the influence that belonging to a particular culture, or simply being human, has in shaping the scientific view of the world.

There is no scientific way to understand the first-person feelings, whims, or imaginings of the scientist (or anyone else) except in the official third-person terms. Since most branches of science suffer physics envy, such terms mean ultimately an account of the motions of ionized molecules within nerves. So, the hard problem is to explain the how the motion of molecules produces the ache of a toothache! Yet, if that sort of explanation is

possible, then scientists' concepts and theories should also be explained ultimately in terms of the motions of molecules! The observable swelling that the dentist sees, and which you can yourself see in the mirror, can be explained in terms of cause and effect. But the painful *feeling* in your tissues is not included in that picture. Nor are your thoughts and other experiences. If a scientific explanation of consciousness is not possible, because of the restrictive definition of science, then what kind of explanation should we look for?

Science presumes that for every effect there must have been a preceding cause, which could involve an infinite regression of the causes of causes. The notion of a *first* or original cause is not defined in science, yet remains an intelligible notion. The original causes we are most familiar with happen to be human beings, beginning with oneself. This notion of *agency*, first acquired by the young child in willing its limbs to move, then extends to other creatures and to hypothetical beings, such as gods and spirits, which are agents with minds. It is precisely this sort of mental agency that science chooses to ignore, which includes the agency of the scientist. No one denies, of course, that scientists are agents with minds, who are responsible for experiments, for presenting their research results, and for publishing papers in journals. But the scientist's agency is not supposed to figure in scientific results. Even for the study consciousness, it is not the scientist's first-person experience that counts, but only facts (assertions) that can be corroborated by other observers. These might include statements made by the human test subject, as in a psychology experiment. But it is the public fact of making the statement that counts as data, rather than its content as personal experience. In principle, objective facts must be accessible to all observers, or at least to select interchangeable observers. In other words: from a third-person point of view. The individual personal experience of the scientist, just as of the test subject, is not relevant because it is not accessible to others. But it is just this experience that poses the problem of consciousness.

One is spared the need to look for an original cause in nature because the scientist is at liberty to *specify* one: the initial condition. The scientist is the original cause! What the initial condition refers to is not some starting point in nature (where there may be no such identifiable starting point), but to a freely specified starting point in some mathematical procedure, such as the boundary conditions of an integration. This works in theory for most scientific purposes, because the intention is to predict the future behavior of a system on the basis of the freely-specified present state of the system as mathematically described. But *how well* it works depends on how accurately the specified input corresponds to reality and how well the mathematical model corresponds to the real system. In other words, it depends on the accuracy of measurements and on the completeness or adequacy of the theory.

What all that means for our purposes here, is that the subjective participation of the scientist in doing science does not appear in the final account of the scientific portrait of nature. If it did, that portrait would reflexively refer to itself, creating an unmanageable hall of mirrors reflecting its own reflections, ad infinitum. From the point of view of a rigorously comprehensive objectivity, omitting that reflexivity is a logical defect; yet, it is a practical necessity. Science muddles through by ignoring the questions it cannot answer and the limits of its own method. Yet, science represents more than a systematic way to deal with the unknown. It also serves as a creation myth, a doctrine to underwrite faith in human powers, a justification to exploit nature through technology. Because it is so deeply empowering for modern society, there is a tendency to kindly overlook its limitations. Yet, the problem of

consciousness confronts us directly with those limitations. Thus, many people think of consciousness as the greatest mystery still unsolved by science (along, perhaps, with the question of why anything exists at all!) Despite the fact that someone perennially comes up with a putatively "scientific" theory of consciousness (as a quantum effect, for example) I hope I have shown that it is not yet even a scientific question. Whether it ever will be, depends a good deal on how the definition of science might expand to include itself as part of the processes it investigates.

The problem of consciousness is to explain the very existence of the first-person from a third-person point of view. However, the "third-person" is a convention of language and of thought, a way of speaking about the transmission of information from one agent to another. An agent (such as a living human body) is not only an object in other people's perceptions, but is also a subject with a first-person view. *All* experience is necessarily first-personal, the experience of an agent. One might claim some personal experience as a fact about the world, but it is that agent's claim, on the basis of *their* experience. We are used to thinking of facts as free-standing truths that exist objectively—that is, independently of anyone asserting them as claims. But this is only another convention. Whether true or false, facts are assertions made by agents—communications from one agent to another, on the basis of first-person experience. Scientific facts and theories are assertions that scientists make to one another and perhaps to a wider public. By trading in *statements*, scientists avoid the problem of explaining or even acknowledging the consciousness behind them. By ignoring the relevance of agents, and of their own purposes as agents, they can focus strictly on inert matter to be manipulated for purposes that need never be mentioned.

An explanation of consciousness, however, must include an account of agents and their purposes, not merely facts about the passive interactions of molecules or other physical forces. In the traditional scientific view, non-living matter is passive in the way that stones appear to be. The modern understanding of cause involves one inert thing impinging on another but not initiating action on its own. Organisms do not fit gracefully into this worldview, and a huge effort in biology has been made to explain the phenomena of life in terms of non-living matter—ultimately in terms of physics and chemistry. This project has been remarkably successful, but at a cost: it consigns the notion of agency as something to explain rather than a principle with which to explain other phenomena—such as consciousness.

An agent is a system that initiates action for its own reasons. By that definition, organisms are agents and stones are not. The behavior of stones is exactly the sort of thing that causal explanation is about, rather than the expression of agents' intentions. Yet it is no more (or less) far-fetched to speculate about the reasons of an agent than it is to speculate about the causes of a physical event. Hume had seriously challenged the notion of cause as a *power* within things to bring about change. His point was that the appearance of causality amounts to no more than an *observed succession* of events. But if there is really no such thing as causal power, why should we reject reason, purpose, or intention in favor of so-called cause? For, in either case, what is observable is a succession of events. Concerning organisms, at least, it may be merely a matter of choice whether to explain a succession of events in terms physical causes or the reasons of an agent—a physical or a logical flow.

If consciousness cannot be scientifically explained, I propose that it can nevertheless be understood in terms of the actions and purposes of an agent. In other words, being conscious is something that certain organisms *do* for a reason. It is not a superfluous addition to outward behavior, but is intimately connected with at least some behavior. It is functional, not epiphenomenal. So, what function does it serve? To understand that, let us look again at the experience of pain and the behaviors associated with it.

Let's say I accidentally touch a very hot surface. My hand may automatically jerk away in reflex. This action happens *before* I feel any pain. But if it doesn't happen quickly enough to avoid tissue damage, I will soon experience a burning sensation, which may linger on for some time. This sensation warns me to protect the damaged tissue. In other words, the response to the stimulus has two components: an initial non-conscious reflex and a subsequent conscious sensation of pain. They have different associated behaviors and serve different purposes. The reflex quickly removes contact with the stimulus. The lingering painful feeling, however is a secondary response. Significantly, *it is internally generated by the brain*, which is why it may persist for the duration of the healing process. It serves not to avoid initial damage from the stimulus, but to avoid future or further damage during healing. That goal cannot be achieved only through the local non-conscious reflex, but must involve a coordinated effort of the entire organism. An organism that could not feel pain would be unable to protect itself from such incidental damage. Protective behavior is associated with pain and depends upon it. (Some insects apparently do not to respond to the loss of a limb or other serious damage with protective behavior, suggesting that they do not feel pain.)

I am using pain as a paradigm example of conscious experience that obviously involves feeling and a particular behavior. However, much of our conscious experience does not seem to involve feeling in the same way, or any associated behavior. In particular, visual experience often does not. But vision and hearing are by definition *distance* senses, so that direct contact with the organism is not involved. Neither, therefore are the responses involved in immediate contact. Because of distance from the stimulus, the organism has time to monitor the environment and consider response on a different level. (Nevertheless, detecting the shape of a tiger in the bush may elicit fear and a flight or freeze response; the color and outline of a piece of fruit may also elicit feelings and prime a creature for action.) The fact that we are largely visual creatures, who are by now to some extent in control of their environment, can give human beings the misleading impression of being detached observers of the world. But feeling remains at the core of this relative detachment.

So, what is feeling? Usually judgment is involved: a feeling is either pleasant or unpleasant. And what is judgment? It is how we evaluate the stimulus, whether it is good for us or bad. We are attracted to what is pleasant (good) and we shun what is unpleasant (bad). That is, the *meaning* to the organism of some sensations lies in the behavior associated with them. In such cases, the connection between the input of the stimulus and the output of the behavior is not a simple causal connection as in the reflex. It is mediated by the organism's evaluation of the significance for it of the stimulus. This evaluation often involves convoluted pathways in the brain. The distance senses allow time for even more convolution.

One may analyze these pathways in causal terms, as though looking at the flow of electricity through a circuit board. But the *logic* of the circuit board is another matter, upon which such causal analysis by itself sheds no light. In the case of the circuit board, which is a human artifact, the logic is not provided by physics but by human agents who designed it.

One would need to know (or guess) what was on their minds. Unless we are creationists, however, we cannot understand the design of the organism by consulting an agent who designed it. The buck must stop with the creature itself—or its kind as a product of evolution. If we want to understand the behavior of the organism, we must try to imagine what it has on *its* mind, so to speak. Of course, that's an act of imagination on our part. But so is the attribution of cause.

The conscious self plays a definite role some of the time in the life of some organisms. This could be likened to the role of a CEO in a corporation. It has limited executive powers and is responsible to the shareholders—the body's cells and organs. Its job is to monitor and coordinate the activities of diverse non-conscious subsystems. Phenomenality is that organism's way of explicitly representing to itself changing conditions, both external and internal. It is the graphic version of an internal language—as the display on a computer monitor is the graphic version of computer code. In the case of the computer, this display is not for the benefit of the computer but for its user or programmer. In the case of phenomenality, the organism *is programmer, computer, and user* all in one. The display is how the CEO keeps track of what is going on in the corporation and wider world.

So, why can't *non*-conscious brain processes do this? Well, to some extent they do—especially in familiar situations that can be handled automatically. That's why we can daydream while driving or zone out while washing dishes. Some people walk (and even drive) in their sleep. There is implicit representation going on non-consciously all the time. But other situations demand conscious attention, which mobilizes resources beyond existing automated programs. Hence, one must pay attention while learning to play a musical instrument or learning a new tune. But after a while you know it "by heart"—which is to say, without the same conscious attention. One could say, then, that the function of consciousness is to learn new behaviours that can become automatic and no longer require consciousness!

It is one thing to understand what purpose phenomenality serves, but we have still not answered the subtler question of what it *is.* The kinds of things that exist in the physical world include things like trees, rocks, clouds, people, animals, chairs, automobiles, molecules, gravitation, etc. While that would include teeth and rose blossoms, there does not seem to be a place on this list for the *ache* of a toothache or the *sensation* of the rose's scent. If phenomenality is not a material thing, then what is it?

I propose that it is an act of the organism communicating with itself, how it represents to itself its own changing state and that of the world. It is a sort of narration or story, like news reporting, or like an animation that is based on reality and constantly updated on the basis of new sensory input. Perhaps the best (and most current) metaphor is *virtual reality*, a simulation produced by a computer program. An animation is usually a fiction created by one agent to entertain another. However, the brain is scripting phenomenality in real time, on the basis of current input, for its *own* use rather than for entertainment or for others. Simulation is often used for serious purposes, as a model of reality that can be fast-forwarded into the future or rewound to the past. Phenomenality has this aspect too. Sensory awareness keeps us apprised of happenings in the real world that can affect us. While phenomenality includes this moment-to-moment experience, it can also project beyond the present moment, since it also includes the (real-time) experience of imagining and remembering. One could

say that the brain creates for itself a useful "show," which is a creative fiction that is continually guided by the world. There is both internal and external input.

I mentioned before that the experience of pain is internally generated, although triggered by an external stimulus. The general truth is that *all* experience, thought, and behavior involves both an internal and an external contribution. Therefore, everything that can enter your consciousness has a subjective and an objective component, which are two "variables" giving rise to experience. High school algebra tells us that an equation with two variables cannot be solved without a second equation in the same variables. Unfortunately, in life there is no second equation! A great conundrum has always been to sort out what in the phenomenal display comes from the object from what comes from the subject. That is, to sort out the real from the imagined or merely subjective.

The senses are not open windows on the world, but more like instruments such as thermometer, motion detector, or smoke alarm. The brain, after all, is perfectly sealed in a windowless chamber, from which there is no exit! Its connection to the external world is via electro-chemical signals it receives and sends out over nerve fibers. Imagine yourself in the analogous situation. You find yourself in an isolation chamber, with nothing but instrument dials and control levers, like in a submarine. Yet, unlike in the submarine, you have never been outside this chamber! These instruments obviously have some purpose, which you must discover through trial and error. But you have no idea yet that there is even such a thing as "outside." You figure this out only as you go along. Through trial and error, you learn how to navigate entirely by instrument, without ever seeing what is "really" out there. *Seeing*, in fact, is no more than your interpretation of instrument readings, as they have been coordinated with control levers. You interpret certain instrument readings as "solid objects." This is like making a relief map of the seabed and underwater world using sonar. Though you are never in direct contact with this world, if the information you have gathered permits you to avoid disaster, then it is "true," and what you are doing is "seeing what is there."

I present this thought experiment as though it involves the learning of a single brain, which is hardly the whole story. Individual brains benefit from the successes and failures of many generations of ancestors. This "submarine" operates mostly on autopilot, controlled by a sophisticated computer whose programming developed through experience accumulated and passed on genetically by thousands of generations of submarines that survived. Through the filter of natural selection, the only submarines that exist are those that have learned to "navigate" in a way that has not resulted in their destruction. Some, but not all, of this they can do automatically.

This brings us back to consciousness as a separate control system, different from non-conscious brain processing. It is as though someone must "be there" to monitor events and take charge when autopilot cannot be relied upon. I certainly do not mean that there is a little person inside the brain, as there is a person in the submarine. On the contrary, the entire "submarine" must in these circumstances act as though it were a person rather than a machine. Indeed, it normally does this so seamlessly as to remain unnoticed. Much of what we have been able to discover about the operations behind consciousness comes from investigating circumstances in which this integration fails in some way or other.

Consciousness *is* that integrated state of "being there." To mix metaphors, it is the virtual reality the brain conjures for its own use as a travel guide to the underwater world

when navigation requires manual control. Thus, everyday experience (phenomenality) is a continually updated internal simulation of what goes on outside the brain. The crucial defect of this metaphor (and its very point!) is that there is no direct contact with what goes on outside the brain. The notion of simulation (like representation) normally implies something already known that it is a simulation or representation *of*. However, this virtual production is not—and cannot be—a literal copy. It is an original creation, though guided indirectly by external input. This is a relationship for which we have no proper metaphor or even an ordinary name—although Kant called the realm of the simulation *phenomenal* and the realm of the inaccessible reality *noumenal*.

What is the nature of this conjuring up of a virtual reality? I call it *fiat*, a Latin word that means decree: declaring something into existence, as in a royal decree or the divine decree, "Let there be light!" It is the creative act of an agent, producing something merely by defining it into existence. It is the act of the author, mathematician, or programmer. The existence produced is not the same as something simply found to be there or to be so (the idea of objective reality). Nor is it the same as *noticing* that something is there or is so (the passive idea of perception). It is rather the idea of actively *making* it so, precisely as specified. If we were speaking of external things, this invention would be sheer magic—conjuring something from nothing, in total defiance of causality and physical law. But we are speaking of phenomenality, of the internal virtual reality, which is not the external reality itself but a co-creation of the mind.

Unlike the natural thing, the conjured thing is exactly what an agent says it is, no more or less. Fictions are of this kind and so are all concepts and all creativity. As opposed to natural things, artificial things are just what they are defined to be. Thus, they are finite and definite in structure, whereas we only guess at the structure and parts of natural things, which we did not make. A simulation is an artificial thing, whereas (as far as we know) the natural reality it simulates is not. Phenomenality is an artificial thing because it is made by the brain, a product of the brain's definitions, so to speak, even when its purpose is to track reality. Being artificial, this virtual reality is definite and unambiguous. This makes sense, because the organism must take decisive actions in order to survive, even in the face of poor or ambiguous information. Hence, those classic ambivalent figures in Gestalt psychology experiments, which can be seen two distinct ways. One does not see them as vague, fuzzy, or indeterminate; on the contrary, perception flips instantly from one definite interpretation to the other, and eventually back.

What we perceive is never a passive reflection of external reality, but always an active creation interposed, for complex reasons that have to do with the needs of the organism as much as with the structure of external reality. What we perceive has a complicated, subtle, and indirect relationship to whatever is "out there." In other words, consciousness is a guided hallucination. It is neither a pure fiction, nor an open window that reveals the world as it "really" is. Instead, it is a co-production, in which the mind's creativity is continually informed by an external reality that can be known only through this co-production itself.

Yet, the naïve realism of our natural senses tells us that the world is *real* and that it is as we see it. While we now know that this appearance is not the literal truth, one can see why such realism is the mind's natural default position—which one philosopher called "animal

faith." Creatures survive by taking appearances at face value, which means to take seriously the power of life and death that external reality holds over their biological existence. Apart from whatever is or is not actually "out there," we are well served by treating as real the world revealed in our experience. This *sense* of realness is an essential aspect of the virtual reality we call consciousness. If it were not convincing in that way, we would not take it seriously; indeed, our species would simply not have survived natural selection. Just as pain must hurt, so the world must appear real. Nevertheless, sages have always reminded us that, however convincing and legitimate, appearances should not be taken at face value.

One lesson that can be drawn from the ideas presented here is that we cannot dismiss the active responsibility we have for our own perceptions and consciousness, no less than for our actions. The causal view of nature—including the brain—is a victim's view, in which things just happen mechanically. The scientific worldview reserves no place for purpose or agents, only for mindless cause and effect. In that view, the world unilaterally determines our experience, and there is little role for personal responsibility. Yet, in its own limited terms, science thus defined has been unable to explain consciousness.

I have attempted to present a different picture, in which the concept of agency plays an essential role in determining both our behavior as organisms and that elusive but most obvious feature of our lives as human organisms: consciousness. The self is not a passive receiver of experience, but a creative force that shares responsibility with the external world. Yet, in this same picture, the self is no more than a feature of the virtual reality the brain conjures for its own use—an *avatar* in the computing sense—representing what I have metaphorically called the CEO. The self is an agent of the body, and perhaps of the species. It has legal and moral responsibilities, but no existence apart from the corporation it serves. This is a very different view of the self from both the spiritual concept of the disembodied soul, which has no motivated basis for agency, and the scientific concept of matter, which simply ignores agency.

In the view presented here, the self and its consciousness disappear when the body ceases to function as an organism. This ending of consciousness is no more (or less) tragic than the ending of the body itself. It is not a separate issue. The self-conscious human mind can scarcely imagine the end of its consciousness and is programmed to fear the death of the body. Like pain, that programming is a condition of being here in life. To some extent, we can bear with pain in the knowledge that it serves the healing of the body; we can also find ways to soothe it. Just so, we can to some extent bear with the knowledge of mortality, understanding that death is nature's prerequisite for the existence of life. We can find compensations to soothe death's sting, which allow us to appreciate our brief existence. In one sense, the whole of human culture is a way to compensate for mortality. Yet, in another sense, culture is the self-conscious human's fullest and characteristic expression of natural agency.