Point of View as an Approach to the Problem of Consciousness

Dan Bruiger
dbruiger@telus.net March 2013
[approx 6000 words]

Abstract

Context Representational and embodiment paradigms, even together, fail to account for first-person *experience*.
Problem Chalmers' "hard problem" signifies not only an unresolved "explanatory gap" but also a methodological gap between diverse approaches such as computation and embodied cognitivism.
Method This paper proposes to bridge both gaps through a consideration of point of view applied to organisms. Internal agency is proposed to accommodate first-person in a manner distinct from phenomenalist approaches, exploring how a cognitive system constructs its experience of qualia through acts of logical fiat, bridging first and third person points of view. This approach attempts to reconcile evolution-based cognitive science with computational representation, by introducing a type of consideration missing from both and applicable to all cognitive systems.
Results The main thesis is that a cognitive system assigns meaning through a process that resembles logico-mathematical supposition ("fiat"). A (third-person) observer can understand the first-person outcome of this process as a psychological completion or projection effect. Moreover, this implies neither neurological isomorphism nor a Cartesian theater.
Implications The argument is intended to encourage a more unified approach to the problem of consciousness and its evolution. A wider implication is that a strict adherence to third-person concepts of agency must be broadened even in the non-biological sciences.

[key words: hard problem of consciousness, agent causation, completion effect, filling in, intentionality, autopoiesis, fiat]

1. Introduction: the problem of consciousness

The very existence of consciousness, or subjective experience, poses a problem to science, which in general deals in third-person descriptions and objectivist accounts. The problem of consciousness is not new, of course, yet seems to remain off limits to a mechanist science. To paraphrase Leibniz, for example, no amount of climbing about in the machinery of the brain explains one's experience of the color green. While there has never been (and perhaps cannot be) a definitive and universally accepted solution, the problem is periodically restated, giving at least an illusion of progress.[1]

There has been genuine progress, of course, in the scientific study of the brain and of cognition. Drawing on computational metaphors, representational theories of mind attempt to explore the organism's role in perceptual processing and cognition, but at the price of considering the mind/brain as an isolated system, and of considering cognition only from a third-person point of view. On the other hand, shunning the notion of internal

[1] What was known fifty years ago typically as "the mind-body problem" is now widely referred to as "the hard problem" of consciousness (Chalmers, 1995); a decade before, it was "the explanatory gap" (Levine, 1983) that needed to be filled.

representation, some modern cognitive approaches—which do emphasize embodied interaction with environment—nevertheless also tend to implicitly behavioral descriptions, and thereby equally fail to come to terms with the nature of experience.[2] Others apply a first-person approach, limited to the human subject and providing only a hand-waving solution to the problem of consciousness.[3] In part, these failures reflect the divergent worldviews of materialism and idealism, and also the differing commitments of a scientific tradition concerned with objectivist description and a phenomenalist tradition concerned with personal experience. There remains not only a problem posed by consciousness itself, but also an ongoing challenge to address it holistically, in a way that reconciles first and third-person points of view, and which steers a course between naïve realism and radical constructivism.

At root of these challenges is the ambiguity between putatively real things in the objective world and subjective experiences *of* them. This ambiguity is arguably an inevitable outcome of reflexive consciousness, to which the subject-object split is endemic. It is not merely a product of language, of European cultural history, or of scientific or philosophic tradition. The very nature of reflexive mind divides thought about itself, making the problem of consciousness doubly hard. It cannot be overcome by a one-sided approach, nor one in which reflexivity or subjectivity play no part. Since cognition and perception can be treated from either a first-person or a third-person point of view, the role of *point of view* must be made explicit. At the least, a reasonably complete theory of mind must hold that cognition and perception involve a joint contribution of organism and environment—not only to behavior but also to subjective experience and meaning.

Scientific narrative is a third-person description that implicitly ignores point of view as a factor relevant to agency. That is, causal agents are considered from the perspective of the scientist, as objects or processes without a point of view of their own. For this reason, there cannot yet be a strictly scientific theory of consciousness capable of bridging the gap between subject and object. At this juncture, no account of an organism's behavior, let alone its inner life, can afford to ignore the process by which the organism establishes its own point of view. Since human beings are organisms, understanding that process is key to solving the mystery posed by consciousness.

2. Point of View

While the problem of consciousness involves reconciling the first-person and the third-person points of view, one must acknowledge that *all* accounts are necessarily first-personal, however expressed. A third-person view of the world, as practiced in scientific description, for example, is ultimately a first-person account by the scientist, camouflaged by the mechanics of grammar. Classical physics notwithstanding, there is no "impersonal" experience, no view from nowhere and no-when.

While a function of language, point of view also reflects the psychological tension of the "explanatory gap" (Levine 1983) between subjective and objective focus, and between the categorically different kinds of objects of such focus. In the scientific worldview, the hard problem is the very existence of first-person experience in the

[2] E.g., O'Regan & Noë 2001
[3] E.g., Varela, 1996

context of third-person description. Strategies in regard to it often reflect an imperative to reductively dispose of this categorical dissonance—for example, by attempting to explain subjective experience in terms of neural activity, in turn reducible to chemistry and physics. Experience is thus typically considered an output of a cognitive system that is implicitly conceived in third-person terms. While this merely sidesteps the dissonance, the problem is not that the explanatory gap cannot be bridged. For, the mind/brain normally *does* bridge it in daily experience! (One could regard this simple fact as proof in principle that the mind-body problem is solvable.) Rather, "the hard problem" (Chalmers, 1996) is hard to the degree that scientists and philosophers are committed to certain kinds of reductive answer.[4] To proceed beyond this impasse requires contextual attention to the ways in which point of view is imposed.

Much of the time, in our awareness the world simply *is*, without reference to point of view. After all, each of us may be described as a biological cognitive system that evolved primarily to deal with the external world, including social reality, and only secondarily to deal with the interiority of self and other, or with 'phenomenal experience'. Even for self-conscious beings, this means experiencing the world primarily as objective, real, and external rather than as an artifact of the nervous system. Such basic realism has obvious survival value, and a version of it is embraced by classical science. Nevertheless, self-conscious beings also have the ability to question the literal truth of their experience, to bracket it as a kind of inner production, for which the self claims some responsibility. In the computer age, this often means viewing one's interior life as a sort of computational output. (For a social creature, this access to one's own and others' interiority may also have survival value.)[5] Thus, we are burdened with two ways of regarding experience—and consequently with point of view, which is built into the grammar of many languages. Accordingly, one tries to understand *how* we come to experience as reality what we alternatively regard as an artifact of the nervous system. While this reconciliation should be possible (given the proof of possibility mentioned above), to accomplish it requires considering the logic and premises of the cognitive system that lead to the conclusive experience of a real external world with its various experienced qualities. One cannot, in search of an explanation, simply recycle the output as the input, without considering the intervening logic, which must be considered from the cognitive system's own point of view, rather than from that imposed by an observer.

3. Embodiment and Representation

---

[4] David Chalmers (Chalmers 1996) famously distinguished the problem of accounting for subjective experience from the "easy" problems of accounting for cognitive behavior: "It is widely agreed that experience arises from a physical basis, but we have no good explanation of why and how it so arises. Why should physical processing give rise to a rich inner life at all?"

[5] Hence, one might say that scientific realism and the 'phenomenological reduction' have common roots in genetic history. Husserl's claim for phenomenology as a new science seems a bit excessive, since its technique appropriates a natural ability. The phenomenological world was hardly unknown to the ancient Vedic culture, for example, nor to the European man in the street at the turn of the twentieth century, who may have lacked a handy name for it. Husserl's formalization marks a phase in the progressive deepening of subjectivity throughout history. Like positivism, it helped redress the ongoing rejection of subjectivity in the sciences, a counterpoint to the natural (realist) standpoint as a tenet of scientific method.

The modern embodied cognition paradigm rightly emphasizes interaction with environment. It acknowledges that neither the behavior nor the subjectivity of an organism can be accounted for by models that consider the nervous system to be self-contained, or to merely process information that is passively received. There is nothing, in accounts that lack the context of embodiment, to show why an abstract and self-contained information processing system should be motivated, have values or directives to govern its behavior, or have a point of view of its own, let alone why it should experience the world as real and external, imbued with phenomenal qualities. Hence, there is nothing to show how "inert" matter can become conscious through self-organization. In particular, the origin of an organism's cognitive premises cannot be accounted for without an appeal to its embodied evolutionary context, which provides the reasons for its reasons. All aspects of organism, including explanations of consciousness, must take into account embodied interaction with the world, which is always implicitly from a point of view.

Yet, there is still something missing in accounts that remain, implicitly or explicitly, third-personal. Following objectivist metaphors, one never quite closes the gap with subjective experience, the first person. Even considering interaction with environment, it is never quite clear how bustling neurons produce the greenness of the color green, or the feelings that imbue values, pains, and pleasures or seem to motivate actions. Even theories that emphasize the organism's active role effectively describe the *behavior* of an integrated *system*, from a third-person point of view, whether this is the organism's molar behavior in regard to an environment or its neural processing.

The mystery remains that there is such a thing as 'first person' at all. What is needed to comprehend this mystery—and lacking in both representational and embodiment approaches—is a concept of internal communication and agency that accommodates point of view.

4. Internal Communication

Embodiment is not simply physicality, but a relationship of an agent with the world, whether established through the present interaction of the individual or the interaction of the kind over generations of natural selection. The behavior of organisms without nervous systems may consist of direct interaction with an environment, without involving internal modelling or even sensory information. Nevertheless, the mediation of some form of internal modelling seems necessary to account for the cognitive functioning of nervous systems, and certainly for phenomenal experience.

'Representation' is here taken as internal modelling. While this does not imply (nor even permit) one-to-one correspondence with external reality, it does suggest internal *communication*. Though a neural mapping may be spatially disposed for reasons of economy, like all communication it is by nature symbolic, and not in principle isomorphic or even graphic. While there is therefore no image in a Cartesian theater (see Sec 8 below), this does not preclude an internal agent that uses the information represented. Internal communication of information, moreover, is *for the organism itself*, not for the benefit of an external observer. The information cannot be divorced from the organism's point of view. The observer is a separate agent, with priorities and a point of

view distinct from those of the organism observed or any sub-agency within it.

Despite the physics concept of information, reification of the genetic "code", and a current trend to treat information as causally effective in biological systems,[6] the organism's internal communications cannot be objectified without losing the sense in which they express the organism's own point of view. A message between human beings may appear to exist objectively, free-standing and self-contained, with a definite information content (as in the case of text or digital files). A third party intercepting such a message is entitled to believe it can be decoded, since it was presumably encoded in the first place. This view of things, however, depends crucially on human intersubjectivity. Its validity cannot be taken for granted when dealing with other creatures and their sub-systems or supra-systems, with the human organism at the sub-personal level, or with nature in general. What is presumed to be *the* information in DNA, for example, is hardly independent of somatic and environmental factors. The extent to which it can be manipulated reliably by human beings, for *their* purposes, depends on fortuitous constancy or control of such factors, even when we don't know fully what those are. In other words, the *organism's* information only happens to have been appropriated by the genetic scientist. Similarly, any "information flow" involved in neural processing is proprietary to the organism.

5. The "Psychologist's Fallacy"

An observer may perceive the organism as adapting to an environment or to its own state. However (or, indeed, whether) an organism perceives its environment, one is tempted to view its actions as taking place upon what *we* perceive to be its environment. It is similarly understandable to identify the structure and functioning of natural agents according to human categories, definitions, and purposes. William James dubbed this innocent foible the "psychologist's fallacy". The very nature of the organism, however, is to be *self*-defining, with its own priorities; it is only incidentally an object of human definition and study. While a machine or other artifact exhibits the priorities of its designers, the organism derives its own through a history of phylogenetic and ontogenetic interactions with environments consisting significantly of other players.[7] This is the embodied basis of cognition, and of the organism's point of view.

If the organism does have sense organs, it responds to changes in these through activity that restores a preferred state, resulting in preferred patterns of sensory input. Even an organism without dedicated sense organs responds to changes of its own chemistry, for example, in ways that either prove adaptive or not. Either way, the organism is, so to speak, flying by instrument. The single cell and the human brain sealed within the cranium face the same challenge: to respond to changes of its state in ways that favor or permit survival. From an observer's point of view, there is a world external to the cell wall, the skin, or the skull. Yet, the challenge for the organism is to survive (to reproduce), which does not necessarily entail modeling an external world. The organism that possesses a nervous system may find it useful to engage in neural activity that the

[6] See, for example (Walker & Davies 2012)

[7] An organism is self-defining, self-assembling, self-maintaining, self-reproducing—in a word: *autopoietic* (Maturana & Varela, 1980). An artifact might self-assemble, self-maintain, even self-replicate, but in the context of present technology it would not be self-defining.

observer takes to represent or model an external world. The observer may conceive such modeling in third-person terms, as a relationship of input to output, correlated by some algorithm. Such a third-person description is all very well until one tries to grasp how one's own representational process (algorithm) comes alive as personal experience imbued with phenomenal qualities—the hard problem. Yet, the situation is no different in considering the experience of other creatures—even other persons—once we admit the possibility of experience at all. One way or another, the goal is to understand the logic of the organism's cognitive self-programming, from its own point of view—and whereby it creates that point of view—when, as observers, we can only occupy our own human and personal point of view.

5. Qualia as "Completion Effects"

The qualities immanent in experience ('qualia') are hardly "epiphenomenal" to cognitive behavior or neural processing. In looking at a source of green light, for instance, one does not perform the action of analyzing a vibration of measurable frequency and then decide to add or project into that "information" the superfluous quality of greenness. Rather, the experience of greenness comes with that information built in; it is an estimation of frequency (among other things)—*in the first person*. Similarly, the perceived quality of a particular musical tone bears information concerning sound frequency, by means of the organism's sensory aparatus rather than laboratory equipment. Sensory qualities are thus not something above and beyond the information they represent, let alone caused by it in the usual sense that separates cause and effect; rather, they are a version of that information.[8] The quality of greenness, in other words, is a way the embodied subject first-personally represents to herself physical properties that might be detected and represented by an observer using laboratory equipment, and describable in physical terms such as 'frequency', 'intensity', 'reflectance', etc. If the specific quality of greenness seems to convey privileged information beyond that involved in the public analysis of light, this is not information about the world per se, but about the embodied relationship of the organism to the world, which includes its evolutionary history.

In third-person description, the observer may regard some neural structure or process (even within her own brain) as a sub-agent (or program) responsible for analysis of some physical variable associated with color perception, for example. Yet, such an observation is itself a first person account by the observer as a molar agent. For the reflexive being, two domains of description are in play: that concerning one's experience as a subject, and that concerning the objects of one's experience, which include the bodies of people and other creatures presumed to be subjects. What is experienced first-personally as the quality of greenness, in the subject's visual cognitive domain, may be assessed as neural response to certain frequencies and amplitudes of electromagnetic radiation, in the domain of description of the scientific observer. Yet, when the observer tries to explain her first-person experience in such third-person terms, the domain of explanation is recycled as its own cause: one appeals to neural behavior to explain one's consciously-experienced perception of that neural behavior, which then is to be explained

[8] Though temporal succession may be involved, as it is in behavioral responses. One could say that an observer, with laboratory equipment faster than the nervous system, has potentially earlier access to the same information as the subject.

in terms of neural behavior... and so on. In the hall of mirrors of this regression, it then appears that "information" (in the observer's cognitive domain) mysteriously gives rise to "qualities" present in the subject's cognitive domain.

Applying the visual and auditory examples above to experience in general, one can now make the general claim that first-person qualities emerge through a process of "filling in" structures that are alternatively mental constructs in a third-person domain.[9] Only in anomalous circumstances do we even notice this process. These circumstances include laboratory studies of perceptual completion phenomena, habituation, adaptations of various sorts, and effects of spatial projection. In such phenomena, features of the subject's experience typically go beyond the sort of events noted by the observer to be justified by sensory input.[10] That is, the mind (first-personally) "fills in" what is (third-personally) "not really there". The contention here is that tones, colors, smells and qualities in general may be understood as such completion or projection effects—as interpolations or projections based on limited "information" (as it may be conceived in a third-person domain).

Yet, one must not imagine perception as a mental coloring book, where a fundamental (scientific) reality is a structural outline arbitrarily filled in with "secondary qualities". If qualia appear to flesh out some "real" scaffolding described by physics, for example, one must bear in mind that such structure was abstracted from first-person experience in the first place. While science traditionally looks for the scaffolding, the point here is to explain the qualia.

Now, in the case of sound, the experience of tone emerges as "wave fronts" impinge too rapidly to be distinguished individually. A single wave front is not perceived as a *tone*; indeed, it is not perceived at all. Rather, a tone is the global effect of a series of wave fronts. On the present theory, it is a quality filled in as a completion effect integrating discrete *neuro-logical* events that are, in turn, responses to physical differences (wave fronts) in the environment. Similarly, the quality of greenness might be filled in between neuro-logical events encoding properties of light. While color perception operates quite differently than that of sound, both visual and auditory experiences are first-personal versions, in an ordinary perceptual domain, of properties or discriminations that may also be described third-personally in a scientific domain of description.[11]

---

[9] I use the term 'fill in' advisedly. Daniel Dennett's (1992) well-known critique of the notion does not bear on the present argument. His essential (and valid) point there is to distinguish between representing a continuous phenomenal field and representing *that* it is continuous.

[10] That is, the external observer. In some perceptual completion phenomena—such as various instances of neon color spreading—the subject is aware both of a discontinuity and its filling-in. In other situations—such as the normal visual blind spot—the subject is not normally aware of either.

[11] The organization of sound perception into octaves is suggestive. Vibrations of frequencies x, 2x, 3x, etc., are perceived as *qualitatively* similar: the "same" note in different registers. The ear responds to a quantitative congruency with an experience of qualitative similarity. An oscilloscope displays the same information: that a frequency and its multiples are commensurate. While the human eye perceives but one octave of the electromagnetic spectrum, it is interesting to speculate that, if the eye were sensitive to a wider range of frequencies, the pattern of perceived colors would repeat in a way similar to harmonics of sound, with qualitative similarity between frequency x and 2x.

As the hyphenated term suggests, neuro-logical events are at once neural and logical.[12] They may be understood either in terms of event causation or agent causation. The pivotal move here is to consider the filling-in process as the action of an agent, from that agent's point of view. The contention is that this process takes place in the way that the quality of being fifty years old is filled in between one's fiftieth and fifty-first birthdays. That is, it takes place through the logical agency characteristic of intentionality—namely, *by simple fiat!*

6. "Fiat": the Intentional or "Semantic" Origin of Experience

The filling-in process is not a matter of event causation, or causal pathways such as encountered in the physics of an electrical circuit and presumably in neural pathways. Rather, it is a matter of *logical* connections or actions, such as encountered in a wiring diagram, flow chart, or computer program, reflecting the intentions of an agent and a point of view potentially different from that of the observer. No strict isomorphism with a supporting physical infrastructure is either required or demonstrable.[13]

However ungainly it may seem to admit agent causation into scientific discourse, if the explanatory gap is to be bridged neural pathways must be viewed as connections invoking the original intentionality of the organism, in contrast to intentionality derived from the observer or programmer. Such connectivity (like logic itself) is essentially conventional, definitional, and symbolic. Internal representation based upon it constitutes what Helmholtz loosely called "perceptual hypothesis". In the present account we take it more literally, as consisting effectively of propositions framed by an agent and amounting to internal communication.

An obvious problem with agent causation, however, is the temptation to project human intentionality into the situation of the organism's cognitive system. One naturally identifies the very sense of intentionality, agency, or proposition—and, certainly, the first person—with human consciousness, will, and language use. However, there is nothing logically necessary about this association. One is not obliged to *define* agency or intentionality in human terms. That we tend to do so is a matter of habit and anthropocentric bias. What is proposed here is not that agent causation in general is an instance of human agency, but the other way around.

To imagine an organism (or some agent within it or encompassing it) formulating "propositions" or "perceptual hypotheses" invites one to substitute one's own consciousness for the point of view of the agent concerned, *as though* one were walking in its shoes. After all, we are used to this empathic identification in the case of other human subjects and some creatures; as children we accorded it to toys. Yet, one is not obliged to accept the invitation at face value. Nor does refusal constrain one to consider a

---

[12] For the use of this term, see Varela…

[13] Because an electrical circuit and its wiring diagram instantiate the same *artifact*, they are functionally equivalent by definition. Yet, one cannot have this assurance about the "circuitry" of the brain, or the structure of any natural thing. In considering neural "mechanisms", whether held to underlie behavior or experience, one is dealing implicitly with conceptual models and possible isomorphism between *them;* how well the model (an artifact) fits the reality is always another question.

cognitive system exclusively from a third-person perspective, as merely an object without a point of view.

With that caveat, one can proceed to view projective completion as the "logical" action of an intentional agent. Such action is equivalent to the mathematician's voluntary act of logical supposition, as in: *"let x represent…"* This posits a premise to be accepted by convention—hence by fiat. Experiences of color, tone, pain, or other sensory qualities "represent" physical properties significant to the organism, effectively through the same constructive and symbolic process by which sounds are made to carry meaning as words, or algebraic symbols gain numerical significance—namely, by convention enforced through usage. Phenomenal qualities that "emerge" in experience are comparable to intelligible meanings that emerge through the babble of spoken syllables, or through the squiggles on a written page. This is an active process, of internal communication largely concerning interaction with an environment full of significance for the organism.

Without invoking a human inner mathematician, a cognitive system may be likened to an applied deductive system—in mathematical terms, an "interpreted" formal system. On the one hand, it is a free-standing, self-contained entity; on the other, it is a system naturally selected because of its utility in the world. In general, those interpretations or applications in nature are valid that lead to survival and genetic success. From an abstract point of view, of course, any convention in itself is arbitrary by definition. Through an agent's intention, however, what is arbitrary (in the third person) is rendered necessary (in the first person), just as mere propositions become truths by embracing them as suppositions or axioms, or by deriving them from already accepted propositions

As a mere symbol system, a cognitive system is gratuitous and self-contained, and the brain that supports it is accordingly viewed as isolated and passive. Just as a language, an algebra, or a computer program can be viewed as an abstract formalism, so a cognitive system has syntactical structure. Yet, a cognitive system is able, like language and mathematics, to refer outside itself to create meaning. Just as a mathematical formal system must be *interpreted* to have meaning or be of use, and a language normally conveys meaning through reference, so cognitive systems come to refer outside themselves through interaction with the world. Such reference is a matter of the organism's agency, in the context of its world, not that of a human programmer or observer. Hence, the computational treatment of representation must be expanded to include the self-programming of the organism, whether by this is meant the individual creature or the kind.

Intentional connection rides, of course, on causal connection—between and within nervous system, body, and world. Yet, because this effectively "semantic" process is not a matter of causal connection, nor about the brain as a syntactical system, experience cannot be explained exclusively in the neurophysiological terms of event causation. Rather, it is about the connectivity of a self-organizing system, meaningful from its own point of view and within its own terms, through which it actively *assigns* reference and meaning to symbols that may appear arbitrary, or have a different meaning, from an observer's point of view. For creatures vitally dependent on an environment, this reference and meaning are established through critical interaction with it. Hence, the brain is not just an abstract symbol processor, like Turing's universal machine, but an instrument of survival. Unlike a computer, the brain is self-defining and self-

programming—through the organism's evolutionary history, through the ontogenetic history of the individual, and through ongoing interaction with the world as a learning process.[14]

7. Evolutionary Significance of Qualia

While qualia in general may be understood in terms of intentional connections, which are acts by fiat of the organism, the significance of specific qualia must be understood in terms of the evolutionary advantages of particular connections and references. The fact that healthy vegetation, for example, is generally experienced as some shade of green is not arbitrary. The sensation of greenness (unlike the word) is not a merely linguistic convention, but a convention of neuro-logical organization, with the force of long genetic precedent. Indeed, the human cognitive system gradually adapts to distorting colored lenses or filters in such a way that verdant foliage—to pursue the example—is restored to a normal experience of greenness.[15] The sensation of greenness, while conventional as a mere symbolic token, is backed by genetic history, just as words come to be imbued with stable meanings through reference to an ongoing world. It is inherently different than the sensation of redness precisely because of the real-world things it refers to in our evolutionary history, from which it cannot be arbitrarily dissociated. This is why 'inverted-spectrum' arguments don't work.

In any language, including mathematics, the meaning of each symbol is intentionally posited. This applies as well to the internal communications of an organism. The greenness of the color green, the hurtfulness of pain, the spaciousness of space, the solidity of objects, even the realness and externality of the world—all these qualities reflect such posited meanings, established by fiat within the organism and conveyed in the "language of the senses".[16] As with the etymology of words, the evolutionary origin of particular intentional connections is a topic inviting study.

8. Internal Agency and the Cartesian Theater

Yet, one may ask, to whom and by whom are such "meanings" conveyed? What are the intra-personal agents involved? These questions ought not to rhetorically dismiss the notion of internal agency, which should rather be taken at face value to indicate an avenue of further research. While the organism as a whole is the agent of its molar behavior, neural processing can only be incompletely understood in terms of efficient cause (that is, without teleology) or following a teleology that belongs to the investigator. It must encompass agency within the organism as well. The computational metaphor avoids this difficulty—by transferring all agency, meaning, intentionality, and point of view implicitly to the human investigator—but at the cost of the explanatory gap.

Descartes was perhaps the first to intuit a computational explanation of mental functioning. His insight into the equivalence of geometry with algebra, of visible shapes

---

[14] While computer programs (thus far) reflect the intentionality (and communication) of human programmers, the logical connections within a living organism are made by the organism itself, by its species, or (in a broad sense) by nature.

[15] See, for example, Neitz et al. 2002.

[16] For one of the few references to "realness" as an experiential quality, see…

with abstract formal operations, must have suggested then, as it does now, a possible avenue toward a "semantic" theory of representation. Just as geometrical figures may be generated by algebraic operations, so might the shapes and colors we experience in vision, for instance, be generated by purely symbolic operations, carried out by neural events in the brain. In any case, Descartes recognized that there is no need for any *resemblance* of such operations to the things they represent, for the relationship is not physical or even structural, but conventional and symbolic.[17] Yet, because of the dualism with which he is associated, philosophers scorn any hint of the 'Cartesian theater', even though Descartes himself points out the logical regression involved in positing observers within observers. While rejecting substance dualism is certainly justified, it should not be used to preclude the notion of an inner workspace, of sub-personal agency, or of internal communication. Only the misuse of point of view ties these concepts to the fallacy of the Cartesian theater.[18]

Clearly there are no literal display screens or iconic representations for an inner personage to monitor, nor any homunculi engaged in text-messaging. There is no inner canvas upon which an inner artist splashes paint; for, as we have noted, the coloring-in does not represent additional information to that represented in neural processes.[19] Yet, the distinction between reality and simulacrum tends to get blurred in the computer age, where simulation plays such a large role in real-seeming entertainments, as well as in scientific research and in public presentation of it. Nevertheless, a computer-generated image contains only what is already potential in the program and initial data that generated it. Such a display is useful to the human user to summarize information and perform further operations upon it. From the point of view of the *computer*, however, such further operations do not require a literal graphic display to be searched, which exists solely for the benefit of the human user. Similarly, a human observer may mistakenly assume that it is useful to an organism to display to itself, in a literal image, some stage in its neural processing.

Yet, such metaphors (of a monitoring screen or control panel in the head, for example) are anthropomorphic caricatures of neural processes that indeed occur. It is not surprising that there may be, in some instances, "something that it is like" for them to be

---

[17] Descartes set out first to reduce physics to the mathematics of space. Through the coordinate system that bears his name, extension is equivalent to number. Physical events in the brain could thus embody logical or even numerical operations, which could then represent spatial relations—and, hence, external reality. Even lacking modern technology, Descartes seemed to intuit what we now call information processing. He recognized that the afferent nerves relay a pattern of *signals* to be interpreted, not a *copy* of the world.

[18] Perhaps the original Cartesian theater was Plato's cave, an apt simile for the epistemic situation of the brain sealed within the skull. Yet, that metaphor underlines the helpless and passive isolation of an imprisoned and immobilized subject. The metaphorical appeal of the computer, as a disembodied and passive information processor, reflects Plato's influence. If computation, thus understood—rather than the active epistemic participation implied by embodiment—became the paradigm of representation, it is perhaps because computers are (so far) not autonomous robots, but passive tools of human users.

[19] As Sartre noted, eidetic or mental images differ from optical images in that they capture only what is already implicitly known concerning the latter. One is at liberty to posit further propositions (creating new information), but there are no further empirical *data* to be gained without further exposure to the original.

occurring. (In fact, one *can* introspectively survey experience in ways that resemble monitoring a display. Every painter, for example, knows how to "flatten" visual space.) The fallacy involved in the Cartesian theater arises from then reifying some object of introspection as a literal image inside the head, with an inner witness to scrutinize it. The usefulness of the notion of internal agency, however, is precisely to understand the logic of sub-personal processing *as though* from a personal viewpoint, but *without* literally substituting the observer's consciousness. Indeed, the rationale in the first place for the notion of *homunculus* was similarly heuristic, as a sort of thought experiment.

9. Wider Implications

Organism is an ancient metaphor that predated machines and mechanism, and which implied a broader notion of causality. While traditions rooted in the religious origins of science, and fascination with the machine, continue to deprecate the immanent reality of nature and endorse exploitive attitudes toward the material world, the scientific view of the world must again embrace an expanded concept of agency, even concerning what is presently considered inert inorganic matter. Self-organization may be key to further advances in many sciences, even in cosmology where it might bear on the resolution of conundrums such as the apparent highly improbable fine-tuning of the universe to life. The *self* in self-organization must in some sense be recognized as an agent and accorded a point of view of its own.

The classical exclusion of teleology from science, with the narrowed focus on efficient cause, has its moral parallel in society: a diminishing sense of responsibility of the self. Even legally, one can now nearly plead innocence on the grounds that "my genes made me do it!" If preoccupation with event causation is an obstacle to scientific progress, even more so does society remain hindered by the sway of determinism and the mechanist metaphor. And if science embraces a broader concept of causation and agency, its influence might encourage a more responsible notion of selfhood.

10. Conclusion

One cannot understand consciousness without considering the internal agency of the organism, as well as its agency in the world, in the context of the evolutionary and ongoing interactions through which internal syntactic systems develop for the sake of semantic reference. The qualia of first-person experience arise "by fiat" within such systems, on an analogy with mathematical supposition.

The existence of a first-person perspective is necessarily problematic when one looks exclusively for an explanation in third-person terms. The very nature of intentionality takes us beyond the traditional science of passive matter, a science that is a product of obsession with the third-person perspective. If it still seems mysterious that there exist examples of matter that can see and feel from a perspective of its own, it is partly because of the simplistic and prejudicial view of matter we retain.

Further Reading

Baars, B. (1983) Conscious contents provide the nervous system with coherent global information. In: Davidson, R. J., Schwartz, G. E., & Shapiro, D. (eds.) Consciousness & self-regulation. Plenum Press, New York

Chalmers, D. (1995) Facing up to the problem of consciousness. Journal of Consciousness Studies 2(3): 200-219

Dennett, D. C. (1992) Filling in versus finding out: a ubiquitous confusion in cognitive science. In: Pick, Van den Broek, Knill (eds.) Cognition, conception, and methodological issues. American Psychological Association [reprint: http://cogprints.org/267/1/fillin.htm#1, Aug 30, 2012]

Dennett, D. C. (1993) Back from the drawing board. In: Dahlbom, B. (ed.) Dennett and his critics. Blackwell, Basil. [Quoted in Pessoa *et al.* (1998), note 9, p747]

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

Pessoa, L., Thompson, E., Noë, A. (1998) Finding out about filling-in: a guide to perceptual completion for visual science and the philosophy of perception. Behavioral and Brain Sciences 21: 723-802

Maturana, H. R. (1978) Biology of language: the epistemology of reality. In: Miller, G. A. & Lenneberg, E. (eds.) Psychology and biology of language and thought: essays in honor of Eric Lenneberg, Academic Press, New York:

Maturana, H. & Varela, F. (1980) Autopoiesis and cognition: the realization of the living. Reidel, Boston

Neitz, J., Carroll, J., Yamauchi, Y., Neitz, M. & Williams, D. R. (2002) Color perception is mediated by a plastic neural mechanism that is adjustable in adults. Neuron 35: 783–792

Oatley, K. (1978) Perceptions and representations: the theoretical bases of brain research. Methuen, London; Free Press, New York

O'Regan, J. K. & Noë, A. (2001) A Sensorimotor account of vision and visual consciousness. Behavioral and Brain Sciences. 24: 939-1031

Varela, F. J., (1994) Autopoiesis and a Biology of Intentionality, In: McMullin. B., Murphy, N. (eds) *Autopoiesis and Perception*, Proceedings of a workshop held at Dublin City University on August 25th & 26th, 1992. Retrieved from http://www.sussex.ac.uk/Users/inmanh/adsys10/as_resources.html on March 27, 2013

Varela, F. J. (1996) Neurophenomenology: a methodological remedy for the hard problem. Journal of Consciousness Studies 3(4): 330-49 [reprint: http://maudsleyphilosophygroup.org/uploads/pdf/Varela_1996_JCS_Neurophenomenology.pdf, Dec 1, 2012]

Walker, S. I. & Davies, P.C.W. (2012) The algorithmic origins of life. arXiv:1207.4803v1. Retrieved from http://arxiv.org/pdf/1207.4803v1.pdf on 1 December 2012