CAN SCIENCE EXPLAIN CONSCIOUSNESS?
Toward a Solution to the 'Hard Problem'

dbruiger [at] gmail [dot] com

ABSTRACT

For diverse reasons, the problem of phenomenal consciousness is persistently challenging. Mental terms are characteristically ambiguous, researchers have philosophical biases, secondary qualities are excluded from objective description, and philosophers love to argue. Adhering to a regime of efficient causes and third-person descriptions, science has no place for subjectivity or teleology. A solution to the "hard problem" of consciousness will require a radical approach: to take the point of view of the cognitive system itself. To facilitate this approach, a concept of agency is introduced along with a different understanding of intentionality. Following this approach reveals that an autopoietic cognitive system can construct phenomenality through acts of fiat. These underlie perceptual completion effects, "filling in," and phenomenology in general. The brain creates phenomenality much as we create meaning in language, through the use of symbols that it assigns meaning in the context of an embodied evolutionary history that is the source of valuation upon which meaning depends. Phenomenality serves as a virtual representation to itself by an executive agent (the conscious self) tasked with monitoring the state of the organism and its environment, planning future action, and coordinating various sub-agencies. Phenomenality is thus not epiphenomenal, but serves a function for higher organisms that is distinct from that of unconscious processing. While a strictly scientific solution to the hard problem is not possible for a science that excludes the subjectivity it seeks to explain, there is hope to at least psychologically bridge the explanatory gulf between mind and matter, and perhaps need for a broader definition of science.

CONTENTS

1.0 The problem of framing the problem

The explanation of consciousness is a major unresolved conundrum of modern science (Crick and Koch, 1998, p105)[1]. A list of the top hundred open questions in the journal *Science* ranks the Mind-body Problem second only to the question of why the universe exists at all (Hoffman, 2008, p87). The challenge for any scientific theory of consciousness is to show how, if at all, it can fit within the materialistic framework. Conceivably it might *not* fit, either because consciousness is not material (or not produced by the brain), or because the framework itself is too constricting. This paper will explore the latter possibility.

The *scientific* question of consciousness carries with it several presuppositions. First, it presumes a point of view aware of itself. Whether or not phenomenal consciousness requires *self*-consciousness, the task to explain it scientifically could only arise for a mind aware of its own activity, for only then could it entertain the category 'consciousness' at all. Second, it presumes a specific line of inquiry, the particular brand of natural philosophy that originated in seventeenth-century Europe, with its special physicalist requirements. Third, what one hopes to explain in material terms is *human*

[1] References (in parentheses) will generally give just the author name(s) and not the date, unless more than one work by the same author(s) is referenced.

consciousness—and (only) by extension that of other creatures and even possibly machines.

Our cultural traditions include the assumption that consciousness is the normal and basal state of human being. Concerning human beings, at least, it is rarely the fact of consciousness that is in doubt, but only how to properly conceive it. Cultural creations, including written philosophical, mathematical, or scientific accounts of nature and of rational thought, are consciously produced. We identify, in other words, with our life of wakeful consciousness, the actions and passions of the conscious self. However, even during the greater part of so-called wakeful consciousness, we may not be in full possession of ourselves (Metzinger, 2013, p1). Since Freud, it has been established that much of human cognition takes place outside conscious awareness. A current philosophical conundrum concerns whether *all* of human cognition could take place unconsciously: whether "zombies"—creatures identical to human beings except lacking consciousness—are a realistic possibility. A complementary question is whether machines could be conscious, and under what conditions. An important issue, therefore, is what functional role consciousness plays in the existence of a creature and in the scheme of evolution. What purpose does it serve, if any? The issue is clouded by the apparent time lag of conscious experience behind neural events correlated with it and on which it apparently depends. If consciousness is effect, can it be cause? How consciousness might influence the brain (or, more broadly, the organism) is the other side of the coin to the question of how it arises from brain states. Is consciousness merely "epiphenomenal," a one-way product of brain activity, or does causality work both ways? This merges with the questions of how intention interfaces with motor action and whether there is free will.

Many proposed "solutions" to the problem of consciousness do no more than restate it, deny it, or somehow explain it away. But, there remains a genuine gap, whether between ontological realities or between the opinions of thinkers. Chalmers (1995) dubbed the persistence of this gap the "hard problem of consciousness." The problem is hardly new, of course, though prior to Descartes the relation between mind and body in Christian culture seemed uncontroversial (Stone, p34). With Descartes, subjective experience became a thing to explain, alongside the things of the external world. True to the mechanism of the age, Leibniz could find nothing in his "mill" whereby to explain a perception. As Thomas Huxley (1866) later put it, "How it is that anything so remarkable as a state of consciousness comes about as a result of irritating nervous tissue, is just as unaccountable as the appearance of Djin when Aladdin rubbed his lamp."

However it has been stated over the ages, at root of the problem is the ambiguity between putatively real things and subjective experiences *of* them. This ambiguity obstructs even agreement about the nature of the problem. It arguably inheres in reflexive consciousness itself, in the subject-object split. 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, in which reflexivity or subjectivity play no part. In that regard, the scientific approach to consciousness may be set up for defeat.

Since cognition can be approached from either a first-person or a third-person point of view, the role of *point of view* enters in and must be made explicit. At the least, a reasonably complete theory of mind must hold that cognition involves a joint contribution of organism and environment to experience and to behavior alike. Scientific narrative is a third-person description that inherently ignores point of view as a factor relevant to causal agency. From the perspective of the scientist, causal agents are considered as objects, or as processes of efficient causation transmitted passively through a system, without a point of view or initiating power of their own. If only for this reason, there cannot yet be a strictly scientific theory of consciousness capable of bridging the gap between subject and object. However, no account of an organism's behavior, let alone its inner life, can afford to ignore its point of view and its agency as a first cause. Since human beings are organisms, understanding that original causality is key to understanding the original intentionality of organisms and to solving the mystery posed to us by our own consciousness.[2]

1.1 Ambiguity

A peculiarity of the problem of consciousness is the very difficulty of properly formulating it. This in itself begs explanation (McGinn, ftn 1, p349). Certainly, part of that difficulty lies in the fact that philosophy, unlike much of science, does not have a precise, mathematically-defined vocabulary. Mental terminology, moreover, is notoriously confusing. The very notion of *consciousness* is ambiguous, if only because, like many other psychological terms, it can be interpreted in a phenomenal or in a behavioral sense. *Awareness*, similarly, can refer either to a subjective experience or to the behavior of taking cognizance of something. Even *experience* bears a dual interpretation, as either a momentary subjective state or a history of events lived through (such as 'work experience' or a 'traumatic experience'). These ambiguities lead to redundancies in expression for the sake of clarity, such as *phenomenal experience* or *conscious awareness,* where a single term ought to suffice. For this reason, I propose the term *phenomenality*, to mean all that can enter conscious awareness. Other ambiguous psychological terms include: mind, thought, attention, perception, cognition, phenomenon, sentience, sensation, contents (of consciousness), sense-data, qualia, representation, intentional (as purpose or as linguistic reference), disposition (which can denote actual behavior or its underlying cause), introspection (as looking within or as a special attitude), mental state (which can refer either to a subject's 1st-person experience or to a 3rd-person assessment by an observer), unconscious (as a temporary or permanent state or as a department of the mind), and even knowledge (which can be propositional or by acquaintance). Related auxiliary terms such as 'information' and 'computation' are no less equivocal.[3] Even a general term such as 'cause' is mildly ambiguous, since it can

---

[2] Historically, three main lines of thought concerning a scientific resolution of the hard problem have been: eliminative, reductive, and emergent materialism (Revonsuo, p17). A successful theory of consciousness would have to answer at least four questions (Gray, p281): (1) How did phenomenality evolve? (2) What is its function and survival value? (3) How does it arise out of brain events? (4) How does it alter the behavior with which it is associated?

[3] While information has a precise engineering definition as negative entropy, in common parlance and in psychology it entails meaning. 'Computation' has more than half a dozen nuances, beginning with the fact that a 'computer' was originally a person! For example: formal

mean anything producing an effect or, more specifically, efficient cause. Similarly, 'reality' can mean one's experience, the external world, or some transcendent abstraction.

The English language (among others including Latin) uses a single word—*to be*—in fundamentally different ways.[4] This ambiguity is at the heart of Descartes' *cogito* ("I think, therefore I am"). In one sense, "I am" means "I exist." But Descartes answers the question "*What* am I?" by asserting that he is a thinking thing. In the context of Christian and Platonic ideas about the soul, this led to the idea that one's consciousness is a kind of substance—an ontological counterpart to the substance of the material world. Whereas his medieval predecessors might have thought that a person's being consists of a body and a soul, after Descartes the person *is* the mind (consciousness), which *has* a body and *has* a soul. The subject stands apart, in stark contrast both to experience and to what the experience is *of*: "I" exist, separately from anything that can enter my consciousness! I am conscious, therefore consciousness is what I am.[5]

Reification plays a pernicious role in thought and in language, which makes little distinction between objects of thought and physical objects. "Object" can mean a part of speech; a recipient of attention, intention, or action; or else a substantial thing.[6] Language conditions us to believe that an experience or possibility must "exist" somewhere—if not in the external world, then in the mind. We are all too ready to conflate kinds of existence, and to make nouns out of other parts of speech.

Some terms name a phenomenon or experience without purporting to explain it.[7] There is a tendency for naming to *substitute* for explanation. *Emergence*, for example, has become a general principle and catchword, but often does little more than designate a process that obviously happens somehow, with unknown details to be taken on faith. Similarly, *epiphenomenon* names a relation without explaining how epiphenomena are produced or exactly why they should have no causal power.

Such ambiguities have been the source of much confusion and dissention in philosophy and cognitive science. The language of analysis concerning mind is frequently beset with awkwardness, talking at cross-purposes, mixed metaphors, category errors, logical inconsistency, and circularity. Circularity is inherent in language, which can only define things in terms of other (known) things, or describe them as *like* other

---

(uninterpreted) symbol manipulation; effective computability; execution of algorithms; operation of digital state machines; information processing; and physically embodied symbol systems (Boden, 2006, p1424].

[4] E.g., it can mean *existence* (to be or not to be); *equivalence* or *identity* (Venus is the evening star, one plus one is two); or *qualification* or *composition* (the lazy dog is brown, God is good, society is people).

[5] Descartes confuses self-awareness, of being in a conscious state, with being consciousness itself. In English, his formula would have more logically read: "I experience, therefore experiences exist." While that syllogism seems trivial, it refrains from jumping to the conclusion that a personal self exists on the evidence of experience. Descartes fails to extend the same skepticism toward evidence for a self as he does to evidence for a world.

[6] Consider these two statements: (1) It is possible that X could own a donkey, and (2) There is a donkey that X could possibly own. Logically they seem equivalent (Martin, p207), but (2) implies the existence of a donkey whereas (1) does not.

[7] Brian Tomasik (2014/2016) puts "qualia" in this category: "So, we feel 'what it's like' and call this 'qualia', and now that we have a name for it, we can think about qualia as something out there. Then we puzzle ourselves about why it doesn't seem derivable from physics."

things that do not require definition.[8] The difficulty of describing the mental as like the physical (i.e, in its terms) *is* the problem posed by consciousness.

1.2 Bias

Leaving aside religious and idealist views—and considerable evidence for the independence of mind from brain—there have been dozens of approaches to the problem of consciousness within the materialist framework, with perhaps hundreds of variants. Some philosophers approach the "explanatory gap" (Levine, 1983) between the physical and the mental by simply accepting dualism as an irreducible fact. But most modern theories involve some form of reduction or emergence (van Gulick, p26). Many propose some version of functionalism. Yet, many proposed solutions fail because they actually address behavior rather than phenomenality. According to Chalmers (1995), however difficult the problems solved, they invariably turn out not to be the "hard" one. On the other hand, not all philosophers agree that there even is a problem to solve.

Some of the latter choose in various ways to dismiss or explain away the apparent need to assimilate consciousness to a scientific outlook. Some seek to reduce the problem to one of semantic misunderstanding; some deny that consciousness plays any meaningful role. I have not found such arguments convincing. Rather, I accept the challenge at face value: to explain phenomenality scientifically, though I don't believe that is possible as science is currently defined. I propose to show *why* it is not and how the current scientific framework can be appropriately modified.

Some theories of representation draw upon computer analogies, essentially by challenging the programmer with the brain's tasks. I think that is on the right track, as far as it goes, since it positions the scientist to consider the organism's proper agency. But this is hardly to say that the brain, let alone the organism, is a digital computer. Nor must it mean endorsing a naively realist perspective, in the sense of presuming cognitive access to the environment as it objectively exists. The computational paradigm has been rightly criticized for presenting the nervous system as disembodied and self-contained—in fact lacking its own agency. Representation has been criticized for simply regressing the problem of perception to an inner observer in a Cartesian theater.[9] As we shall see, the problem, precisely, is to make the leap from "agency" (a 3rd-person concept) to "experience" (in the 1st person). And any approach whatever must eventually come to terms with the paradoxical issue of access to the world-in-itself, first addressed by Kant.

Unfortunately, while acknowledging that paradox, proposals that incorporate the action and embodiment of the organism typically do not make the required explanatory leap. In the end, they explain behavior, but not experience. The sensorimotor theory of O'Regan & Noe (2001), for instance, is well motivated by the commonsense view that the brain does not operate in isolation and that perception involves the active

---

[8] It is fundamentally challenging to define mental states in a way that does not circularly refer to other mental states. Attempting to define even simple psychological terms can end in proliferating mental terminology. E.g.: "A goal can be conceptualised as a representation of a possible state-of-affairs towards which the agent has a motivational attitude. A motivational attitude is a kind of 'propositional attitude' (Beaudoin, p41)."

[9] A trap that Descartes himself recognized.

participation of the organism. Its radical claim seems to be that there need be no neural correlates of consciousness, no isomorphism between neural activity and experience, indeed no representation at all.[10] In the end, however, it deals with behavior and not phenomenality. While the nature, role, and extent of internal representation remain topics of controversy, I will simply assume here that some form of symbolic representation—if only memory—is necessary, if not sufficient, to explain both phenomenal experience *and* the behavior of higher organisms. This tenet is grounded in the fundamental principle that all cognition is co-determined by both the organism and its world, by subject and object conjointly.

Disregard of this fundamental principle leads to various biases, not to mention absurdities, and some investigators side more with external and some more with internal factors. This is perhaps understandable, since the situation resembles an equation in two variables, for which there can be no solution without a second "equation." Failing that (as it must), one "variable" must be arbitrarily held constant to explore the other. Scientific experiments, for example, are designed to artificially isolate a single causal factor. The very idea of objective reality involves focus "out there," on the object, without regard for the subject. Similarly, focus "in here" on phenomenal experience may disregard the object, leading to pure idealism or solipsism. Given the principle of co-determination, the real question may be why human thought is drawn toward such mutually exclusive extremes.

All forms of cognition—including science itself—involve the interaction of subject with world. We must find a way forward that excludes both naïve realism and naïve idealism. Fortunately, there seems to be growing recognition of this common-sense truth (e.g., Varela et al; Thompson et al; Mausfeld). Not only are behavior and phenomenality co-determined by the organism and by the world conjointly, but organism and world also affect each other objectively in various ways. The organism has the capacity to physically shape its environment, as well to define what is salient within its sensory input and respond appropriately. It adapts the world to itself and itself to the world. These are distinct capacities; one is the result of the organism's motor action upon the world; the other is a result of its cognitive action, which may include information gained through motor action and which affects primarily the organism itself and only indirectly the world.

The dilemma of how to handle this co-determination spawns a diversity of approaches toward the problem of consciousness. Classically, some reject the problem as a category mistake (Ryle) or as an ill-considered side effect of dualism (Ayer, p124). Some disclaim the explanatory gap as an illusion (Loar, sec 8); some, as a defect in our neurophysiology (Jack, 2013) that disposes us to see an explanatory gap where there is none. Some defer the problem in favor of the "easier" tasks of finding neural correlates of consciousness (Crick & Koch, 2003). Others (such as Leibniz and Feigl) take refuge in psychophysical parallelism or a double-aspect theory. The latter approach might accept *that* there are two "aspects" without explaining why (Humphrey, 2000b, p6), or else may simply translate an ontological dualism into other terms, effectively doing no more than to restate the problem. Still others resort to panpsychism (Chalmers,1995), to the quantum microphysics of the brain (Hameroff & Penrose, 2014), or to blatant idealism

[10] O'Regan & Noe, Sec 9: "The solution to the puzzle of understanding how consciousness arises in the brain is to realize that consciousness does not in fact arise in the brain!"

(Hoffman, 2008). Although behaviorism is now dead as an official doctrine, the troubles that led to it have not been resolved (Price & Aydede, Sec2). Some philosophers (Dennett, 1991) and many scientists still prefer some version of behaviorism as the only approach compatible with science as it stands, even when it means ignoring phenomenality. In short, there has been a great deal of wandering in the wilderness, where mind remains untamed by common sense let alone by physics—a situation exacerbated by confusing terminology, contentiousness, and talking at cross purposes. Mind remains the elephant in the room for science, and the diverse approaches to it suggest the story of the blind men who each describe a different part of the creature and can reach no consensus. Indeed, perhaps one reason for defining cognitive science in formal terms, of formal representation and symbol manipulation, is that it is at least possible then to know just what one is talking about (Boden 2006, p1417)!

1.3 The present approach

Apart from contextual material, the goal of this paper is to explain phenomenality, and to show why an explanation in strictly scientific terms is not possible. I do not argue for physicalism, but assume that some appropriately modified version of it will be compatible with a satisfying explanation of phenomenality. (Justifying physicalism is a separate task.) The problem addressed here is to explain how a physical system (the brain-body-environment) can produce what we call conscious experience, or phenomenality. Scientific concepts and practices may need to be modified or expanded toward that end.

Similarly, the goal here is not to defend internal representation, qualia, the computational metaphor, constructivism, etc.—all of which I take to be useful in context. Nor is it to classify, review, or criticize various approaches; to take sides in any of a number of current debates in philosophy of mind; or to engage in hairsplitting over technical notions. Much less will I try to *evade* the problems posed by consciousness, through sophistry or in the name of some proper way of speaking. Rather I accept Chalmers' (1995) challenge to find a "psychophysical theory" that explains how physical processes give rise to personal experience. Accordingly, I will draw upon aspects of other's work that support my thesis without dwelling on aspects that don't.

I will propose that what is needed to comprehend the mystery of experience is a concept of internal agency and internal communication that accommodates the point of view of the organism itself as the constructor of both its experience and behavior. While opinions differ on the relative significance of key factors, a reasonable theory of mind must hold that cognition involves an interaction in which organism and environment jointly contribute to subjective experience and meaning as well as to behavior. This requires an approach that somehow reconciles first and third person points of view; which acknowledges and then naturalizes the first-person perspective (Frith & Metzinger). If such an approach can be found, it might be possible to steer a course between extremes that exemplify the divisiveness which plagues philosophy of mind, reflecting the subject-object split itself.

Expressions such as *conscious awareness*, *conscious experience*, and *phenomenal experience* are sometimes used here for variety and nuance, though generally I prefer the term *phenomenality*.

2.0 What makes the hard problem so hard?

At core, the problem posed to science by phenomenality is a child of the abstractions of metaphysical thought (Robbins, p176). Historically, what is meant by *the physical* excludes phenomenality by definition. One aspect of the problem concerns the ontological status of consciousness in the physical world. But, it may be possible to show how phenomenality is physical in origin without requiring it to fit within a physicalist ontology. After all, physics itself is replete with inconsistencies and ontologically dubious entities. Physicalism holds that all truths about phenomenal consciousness supervene on a physical description of the world. However, even apart from the issue of consciousness, a complete physical description of natural reality may not be feasible (Bruiger, 2016). In a sense, physicalism is the problem rather than the solution, since our cultural and historical faith that the world is fundamentally physical leaves no place for subjectivity and consciousness (Crane & Patterson, p1). In addition, several specific conditions or issues render the hard problem hard.

2.1 Secondary qualities

The physical sciences excluded what became known during the Scientific Revolution as secondary qualities: those aspects of sensory experience that do not unequivocally reflect properties of the external world, especially those qualitative aspects that do not lend themselves to mathematical treatment. In the terms of the basic framework of co-determination, this exclusion is a result of focusing on the 'object' factor in experience by ignoring the 'subject' factor. Another way to state this is that science normally takes a third-person perspective on its objects of study, thereby excluding first-person accounts. It deals in descriptions rather than experiences. Only entities and processes that can be approached and described third-personally qualify to be considered physically "real" (Revonsuo, p41). But conscious experience is by definition first-personal and very real to conscious subjects. The problem posed to science is then why there is such a thing as the 'first person' at all (MacLennan, Sec IIIA).

Secondary qualities (such as color, sound, taste, smell, and tactile feel) are ambiguous insofar as they seem to depend on both the external world and the perceiving organism. They are excluded from the ontology of physics because of this ambiguity, since they entangle the object with the subject. However, according to the principle of co-determination, so-called primary qualities too must also be functions of both subject and object, similarly entangled. Co-determination notwithstanding, the problem is the different *status* of primary and secondary qualities. One difference is that the dependence of primary qualities on the subject may be factored out from their dependence on the object (for example, position may be made explicit with regard to a frame of reference, so that it does not depend directly on the observer). Secondly, being extensional, they may be more easily quantified and measured. Primary qualities (position, size, shape, structure, motion, etc.) are mainly accessible through the visual sense. Treating them as *quantities* facilitates prediction and control. While vision is associated with rationality and the ideal of objective truth, the "lower" senses are often keener in animals and thus disdainfully associated with animality, appetite, and lower functions.

Secondary qualities are banished from physical description on the grounds that they are subjective artifacts of the organism's perceptual processing. (Again, given co-determination, *all* qualities are artifacts of perceptual processing, which may *also* reflect objective properties.) While the objectivity of science relies heavily upon the visual sense, vision includes color perception, so that subjective quality re-enters science by the back door. The question of whether external objects "really" have color (that is, whether color is a property residing in the world or in sensory organ and brain) should be compared to the question of whether odoriferous substances "really" smell. Detecting odors clearly is an ability of the organism to discriminate the real presence of airborne chemicals. The subjective experience of a smell refers to an objective reality. Similarly, color experience refers to objective properties (wavelength, reflectance, ecological significance, etc.). Hence, "color", like "odor," can refer equivocally to subjective experience (qualia), to a capacity of the nervous system (discrimination), or to a property of the external world (wavelength, etc.). Given co-determination, this ambiguity is to be expected. What is more curious is that some aspects of experience have been earmarked as objective.

Thus, one reason for the intractability of the hard problem is the tradition in which "secondary" qualities are artificially separated from "primary" ones, effectively on the basis of special properties of the visual sense. Senses too closely dependent on the body are not considered a source of reliable knowledge. For Descartes, they were not sources of knowledge at all, but mere phantasm. The modern version of this dismissal is the notion that qualia lack any intentional or propositional content. The Cartesian position is that ideas of pure mind (such as mathematics) have a tautological certainty that can be formulated as propositions, whereas the "phantasms" of sensation are unreliable in part because they lack propositional content (Cottingham, p136). Against common sense, Descartes brackets "sensible qualities" as mere appearances, describing his own bodily sensations in an impersonal manner that disowns any cognitive interpretation assigned to them. In modern terms: structure and function alone are of scientific interest. However, qualia *do* convey information about structure and function—not only concerning the external world, but also concerning the internal states of our bodies and the *relationship* between our bodies and the world. Part of the problem, therefore, is that such information plays little role in modern science (Jack et al, 2007, p12).[11] The other side of this coin is that the primary qualities, too, depend on the organism. The problem posed by phenomenality concerns how the mind constructs *all* aspects of experience, including shape, number, position, size, etc.

The distinction between primary and secondary qualities, implicit for the Greek atomists, was made explicit by Galileo, Boyle, and Locke. Leibniz, Berkeley, and Hume, however, questioned it, thinking that primary qualities must also be mind-dependent (Edwards & Wilcox, sec2). Kant began the project of showing how we mentally

---

[11] A science in which it *did* play a significant role would take into account its own cognitive states, biases, and epistemology. See Bruiger (2014).

construct them. It remains to clarify how the primary qualities, too, reflect our animal needs—for instance, to what extent perception of space depends on motor action and capacities.[12]

In contrast to primary qualities, Galileo writes in *Il Saggiotore* that "these tastes, odours, colours, etc… but hold their residence in the sensitive body; so that if the animal were removed, every such quality would be abolished and annihilated." This distinction hinges on that between the "sensitive body" and the "rational soul," which became the distinction between animal and human. It is this bias, no doubt, which prevents Galileo and Descartes, unlike Kant, from imagining that removing the animal would abolish the primary qualities as well! While the world as perceived by the mere animal would disappear with its corporeal sentience, the consciousness of man is presumed to continue after the death of the body and to reflect the consciousness of God, through which all things are maintained and known.

Ultimately, the primary qualities are supposed to represent observer-independence: how nature is carved at its real joints, so to speak. However, someone is required to do the carving, so there remains a subjective factor involved in putative objectivity. Moreover, unlike physics concepts based on spatial extension, concepts of mass and force refer not only to vision but also to touch and proprioception—which reside in the "sensitive body." The senses are body-centric, detect differences only, and serve survival. In contrast, scientific instrumentation is supposed to measure differences independently of the organism and thus better serve the ideal of objectivity. The organism is tuned to features of the mesoscopic world, not to deeper features independent of its own scale (Singer). However, the ideal of objectivity may itself be a survival strategy. The relative simplicity (Turok) at both extremes of scale, as compared to the manifest complexity of the mesoscopic world, could turn out to be illusory, due to an anthropocentric bias.

It may appear that phenomenal qualities (qualia) are "ineffable," performing no function in and of themselves.[13] Ineffability does not imply vagueness or vacuity, however, only inability to form propositions and concepts based upon them. The distinction between propositional knowledge and knowledge by acquaintance corresponds to that between primary and secondary qualities. Often different epistemic agents are involved, but even when observer and subject are one and the same, they act in different capacities. Knowledge by acquaintance is first-personal; propositional knowledge is third-personal.[14] Propositions are distilled, with information loss, from

---

[12] Cf. Singh & Hoffman, sec1: "Because natural selection has tuned our perception of shape to be an adaptive guide to behavior, our perception of shape has evolved to be tightly coupled with our actions… Thus the detailed properties of perceived shapes, such as their symmetries and parts, are not depictions of the true properties of shapes in an objective world, but simply guides to adaptive action." Cf. also Dehaene: "My claim is that number is very much like color. Because we live in a world full of discrete and movable objects, it is very useful for us to be able to extract number… This is why evolution has endowed our brains and those of many animal species with simple numerical mechanisms."

[13] Lamme, sec 1: "Qualia are defined as ineffable aspects of information: the redness of red, stripped of every possible functional property or reactive disposition."

[14] Of course, propositions are asserted *by* an agent, and in that sense are first-personal. However, 'I see that the tomato over there is red' is a different proposition than 'That tomato is red'. The first claims responsibility for the perception, the second does not.

phenomenal experience, which cannot be reconstituted solely from the former (Talbott; Boltuc & Boltuc, p272).[15]

2.2 The mechanical philosophy

The exclusion of secondary qualities is not the only reason why the hard problem is hard. A general mechanical style of thought permeated science from its inception, which embraced mechanism, determinism, reductionism, deductionism, and eventually other "isms" such as physicalism, epiphenomenalism, and behaviorism. To paraphrase Leibniz, no amount of climbing about in the machinery of the brain explains one's experience of sensations, for example, of color. Unlike Leibniz, Newton (1730) failed to see any dilemma posed by phenomenality: "colours in the object are no thing but a disposition to reflect this or that sort of rays more copiously than the rest… to propagate this or that motion in the sensorium…" The success of Newtonianism set the tone to ignore the scientific study of consciousness until late in the twentieth century. By that time the computer had provided both a metaphor and a tool for understanding mind as an information processing system. However, there was still little recognition that the methods of physical science as it had been defined were insufficient to determine even what an information processing system actually does (Sloman & Chrisley, 2005). Hope lay rather in trying to replicate the system's behavior, shifting from a descriptive approach to a prescriptive one (i.e., programming).[16]

The notion of epiphenomenalism—the causal ineffectiveness of phenomenality—is a byproduct of a simplistic and mechanistic view of single causes preceding effects in time. Epiphenomenalism implies that the conscious self is simply "along for the ride," playing no causal role but passively experiencing effects that have already been determined by events to which there is no conscious access. However, organisms involve circular and multiple causation, operating on many levels. Let us propose that the role of the conscious self is precisely to respond, after the fact, to preceding events in the nervous system. This hardly negates a role for the conscious self to play, any more than the operations in a computer negate a role for the user to play. Quite the contrary,

---

[15] Cf. also Revonsuo, p41: "In NaCl (common salt), there is nothing constituting the 'salty' quality that we taste in our phenomenal consciousness." The "real" properties of salt—its chemical structure and interactions with other substances, as opposed to its salty taste in the mouth—are supposedly those that depend in no way on the subject. But in truth there are no such properties, only properties that depend on the subject in widely varying ways. Taste depends on taste receptors and color depends in a complicated way on receptors in the retina. So-called real properties are theoretical abstractions based on information gleaned via other senses. Molecules of Na and Cl are also abstractions, which we *envision* as having shape and spatial disposition. Since NaCl is a theoretical construct distilled from diverse percepts, the percepts (qualia) on which it is based cannot be reconstituted from it. If we wish to explain the salty taste and even appearance of NaCl, we must consider also the physiology and intentionality of the perceiving organism.

[16] If you had designed the information processing system in the first place, it should be clear what it is doing. But if you came across an unknown system, you might try to reverse engineer it as well as perform experiments upon it, take it apart, etc. In the case of organisms, we are dealing with systems that engineered themselves—a notion foreign to classical science.

consciousness may serve as a user interface for making high-level decisions that *do* have causal effects on the operations of the system.

The general view of causality is skewed because of focus in physics on simple isolated systems and efficient causes. This is reinforced by focus on physics itself as "fundamental," as the view of the world as it "really" is. This, in turn, is a result of cultural choice or historical accident more than logical necessity. The dominance of mathematical physics trains us to look for single causes, which allow for clear and predictable linear relationships. This often works well, of course, in the inorganic world, where one cause stands out and others can be neglected as background. But single-factor analysis leads to trouble when applied to the living world—and to phenomenal experience, which is a joint product of organism and world, neither of which can be neglected as a background for the other.

Causation in general has simply come to mean *efficient* causation, which was intended to deal with non-living systems and forces. While neural cells may be viewed as passive *transmitters* of processes of efficient (i.e., upward or physical) causation, they may also be viewed as active *initiators* of signal transmission. The brain and its parts may be viewed as *agents*, involved in *final* (i.e., downward or mental) causation (Eady, p17-18). Just as there may be multiple efficient causes, we are not in a position to specify the "*right* answer to the question: What is it for?" (Dennett, 1987/1998, p319) 'Information' within organisms must not be construed only in the strict thermodynamic sense; it implies a sender and a receiver—agents with purposes, initiating processes in various directions within the organism. The problem of mental causation results from the original exclusion from physics of such notions as agency, teleology, final and multiple cause.

Thomas Huxley's (1874, p240) famous example of a train whistle does not aptly illustrate epiphenomenalism. For, only when the system concerned is the isolated engine does the whistle appear to be purely effect and never cause. The whistle *does* have causal effect when the system includes the people for whom the whistle is intended as a signal. The whistle is functional in the overall operation of the rail system. The same is true of consciousness. It may seem to be the passive byproduct of brain processes that would produce the relevant behavior without it. But, in fact, some behavior can only occur consciously, playing a high-level causal role in the overall management of the body. While this does not *explain* the consciousness, it obviates its apparent uselessness.

At face value, the experimental results of Libet (Libet et al.)—that the experience of willing a movement occurs *after* neural processes leading to the motor action—seem to imply that there is no free will, only the deterministic activity of the nervous system. But, this would be to misunderstand both determinism and neural activity. The fact that the conscious self experiences the willing of an action after it has already been unconsciously initiated suggests rather that the role of consciousness is precisely to monitor processes already under way.

2.3 The problem of cognitive domains

Attempts to explain how the mind builds its picture of the external world usually begin with the very picture of the world they attempt to explain. I call this circularity the 'problem of cognitive domains.' A *domain* is generally a set of elements upon which some operation is to be performed, such as a mathematical function or mapping. The problem of cognitive domains is the dilemma of circularity that arises when the domain that is the output of a cognitive process is recycled as its own input. This occurs, for instance, when the physical world that appears phenomenally is presupposed as a point of departure in order to explain its own phenomenal appearance. The output of mental processing is recycled as the input. A simple example is how three-dimensional (depth) perception is reconstructed by the brain on the basis of two-dimensional information projected onto sensory surfaces; however, these surfaces exist in a world that is already presumed to be three-dimensional.

This general epistemic dilemma applies not only to ordinary perception, but also to science as a form of cognition. The scientific strategy in regard to the problem of consciousness is to explain the contents of consciousness ultimately in terms that originate in the world as conceived by physics; however, *that* world is but another content of consciousness originating through the very processes to be explained. The presumed point of explanatory departure may include such elements as photons, neurons and electrochemical processes. Through these physical elements the brain builds up its concepts of those physical elements: photons, neurons and electrochemical processes, etc. In short, we treat our image of the world as though it was the world itself, which is then recycled as the basis upon which that image is constructed!

The problem of cognitive domains concerns the relationship between map and territory, when access to the territory is afforded only by the map. It invites a kind of epistemological cheating (Kenny), an attempt to "peer around the edges of our 'constructivist goggles' to sneak a look at the 'real thing'…" Several authors have acknowledged the problem this poses for the study of consciousness, and also for science in general.[17] It results, for example, in the presumption that the world as humans perceive and conceive it is the pre-specified domain to which the target organism responds. If human cognition is merely one possible form and not a god's-eye view, then it is a mistake to take it as a standard acted upon by other creatures, or that their perceptual world is a subset of ours. Nor is the physicist's world the ordinary world of human cognition.[18]

Computation mechanizes logic and other more informal thought processes occurring in the brain; to then explain such processes in terms of computation (within the brain) is circular reasoning. Computation is tacitly defined in terms of notions such as representation and semantics, while those very notions are widely supposed to be explainable in terms of computation (Smith, p2). On the other hand, to posit that the organism deals only with its own inputs and responses does not avoid the question of a real world external to it. For, quite apart from what lies beyond its boundary, the organism itself at least is presupposed to be real, existing in a world that happens to include the investigating scientist. Neither a naïve realism, nor a constructivism that is implicitly solipsistic, are coherent approaches that avoid paradox. Epiphenomenalism presents the paradox that science, as a manifestation of mind, must itself be

---

[17] See, e.g., Shepard & Hut, p143-4; Hoffman (2009, sec 1.4); Mausfeld, sec3; Varela et al, p13.

[18] Moreover, taking the physicist's world as a starting point for causal description of psychological processes only works in restricted realms, such as the neurophysiology of receptors and reflexes.

epiphenomenal: scientific arguments and truths must be considered reducible to causal processes in the world and in the brains of scientists (Jonas, p129; Varela et al, p11)!

2.4 The privileged visual sense

We have already noted the special role the visual sense plays in objectification. For one thing, it is key to the distinction between sensation and perception: one feels a pain as a sensation, but perceives a tree (Reid, sec 6.20). Especially for vision, it is challenging to isolate sensation from the perception to which it is integral. Although visual space can be "flattened" and colors and shapes attended to, it is uncommon to speak of visual sensations.

Supposedly, I have unique access to my sensations, but not to my perceptions, which are effectively assertions about the public world: "Any number of people can perceive the nail sticking into my toe—but only I *feel* it" (de Quincey, p77).[19] As a distance sense, vision is detached from contact with the body; and its freedom from immediate implication for the organism allows visual percepts to be the basis of concepts (Jonas, p31). Like photons, visual images bear no significant *physical* impact on the organism. The visual field surveys at once many objects of potential encounter, whereas other senses may reveal them sequentially (Jonas, p145). Visual experience *seems* relatively divorced from affect, motivation, and immediate response—in other words, from the intentionality of the organism. It strives to objectively represent the *world*, rather than the subjective relationship of the organism *to* the world. The detachment of visual sense serves as paradigm for the ideals of rationality, objectivity, and truth.[20] Accordingly, the problem of mind is preeminently a problem posed by the visual sense as compared to the other senses. It seems one must turn to other senses to recover the organism-centered intentionality at the base of perception.

However, the one quality of visual perception that seems ambiguously both subjective and objective is color, and color perception has been notoriously troublesome and pivotal in various arguments about phenomenality. Because color is a paradigmatic sort of quale, the confusions around color perception are one reason the hard problem has remained hard.

---

[19] Of course, this is because my nervous system alone is hooked up to my big toe, so only I can feel the pain there, while many people's nervous systems can visually access my toe. However, only *my* nervous system is hooked up to *my* eyes, so only I can have the *particular* experience of seeing the nail in the toe from just *this* perspective in space and in just *this* precise way. In that sense, my visual "sensations" are as proprietary as my bodily sensations.

[20] Of course, neither visual perception nor the concepts of science reveal the thing-in-itself. Cf. Solms, p174: "…things in themselves (as opposed to our perceptions of things) do not possess visual quality. Vision is a property of the subject…We tend to think of objects as if they consist ontologically in the forms by which they are presented to us visually. We think of our visual image of the world as if it were, in a word, *reality*…" As we shall see, the problem presented by the seeming objectivity of the visual sense is to understand how it nevertheless derives indirectly from affect.

2.5 Limits of cognitive access

The peculiar nature of the epistemic situation of an organism is that it must deal with the external world in order to survive, but has perceptual access to that world only by means of its own physiology and internal processes. There is no "direct" access to the "world-in-itself" that bypasses its cognition. The human brain, for example, is sealed within the skull, with external senses that do not provide open portals on the world but only a stream of data to which it must respond in a manner that allows the body to act in such a way that it survives. One can say, for shorthand, that the brain's model of the external world must be realistic, but there is no way to get outside the skull to compare it to reality. One cannot stand outside oneself to assess the degree of fit that one's model bears to the world (Sterpetti, sec 1). The only feasible measure of this fit is whether it favors survival, which can hardly be known in advance. Moreover, survival depends on other factors as well. Like the concepts of reality and objectivity, the notion of accessing the "thing-in-itself" is paradoxical, since it presumes impossibly to isolate the object from the subject. At core, this *is* the mind-body dilemma.

After Plato, Kant seems the first to have fully appreciated this dilemma. One could say that physics is our best shot at describing that absolute world behind the world of appearances, but physics is no more exempt from this fundamental dilemma than ordinary cognition. Science provides alternatives to our perceptual models, but these are still no more than models; it shifts the stream of data to instrumentation, but the burden of interpretation remains. There is still no access to how the world "really" is apart from interactions with instruments and scientific theorizing. We may say that scientific cognition is more accurate than ordinary cognition, where accuracy means the ability to measure and predict; but the ultimate test for both remains viability. Despite the proliferation of human civilization on the planet, science is too recent on the evolutionary scene to conclude definitively that it enhances our prospects.

The attempt to achieve an absolute portrayal of the world—or at one least more reliable than that provided by the senses—began in European traditions with the speculations of the Greeks. The flowering of this project, eventually in modern physics, meant a split between the world as ordinarily experienced and as portrayed in science (Mausfeld, 2002, sec2, Example 2). Despite Kant's admonitions,[21] we tend to defer to science for a vision of the world as it truly is, in contrast to the merely "subjective" world of perception. This practice underwrites the problem posed by phenomenality to the scientific worldview. Why reduce phenomenality to physics unless you believe that physics is the true—or at least the more legitimate—account of the world?

The organism's phenomenal experience may be likened to a user interface between the organism and its environment (Hoffman, 2009, sec1.5). In those terms, the organism has direct access only to this interface, not to its environment. Nevertheless, it is functional for it to behave as though it were dealing directly with that environment and not with its own constructs. In fact, we do not experience the world as taking place inside

---

[21] See Iglowitz (2010, p270-71): "While fully affirming the *existence* of our external world… as a necessary prerequisite to reason, Kant concluded that we are inherently incapable of knowing any of its independent properties, (to include time, space, extension, tactility, impenetrability ...), that is, we are incapable of knowing them *independently of* their revelation in, and in combination with, human cognitive forms…"

the skull but as “out there” in real external space. We are unaware of how this projection takes place; the internal processes responsible for perception do not generally present themselves to our consciousness. There is no biological need or evolutionary justification for the organism to keep track of its own neural processing, which would be a cognitive burden. It is only the final output that counts.

This effect is often called *transparency* (Metzinger, 2005, sec 1.1). We see the world *through* our perceptual processes, which themselves remain invisible. This is not a logical but a biological condition. Yet, it is no coincidence; for, the nervous system is organized precisely to provide this illusion of a transparent window. In general, the organism has no need to access its internal processing, and doing so in many cases would be counterproductive. This has the consequence that the qualitative states of experience (qualia) have a gestalt and ineffable aspect; they are difficult to analyze introspectively, since we lack cognitive access to the processes that construct them (O’Regan 2010, p11). It also means that we do not normally question our perceptions, a fact that has social consequences and is the basis of naïve realism. In compensation, it motivates the ideal of objectivity and skepticism concerning the senses; for, skepticism implies the possibility of transcendent knowledge (Talbott, 2004). In short, transparency and the outward focus of the organism have rendered the mind’s understanding of itself problematic from the outset.

2.6 Psychological conditioning

Lastly, various habits of thought, and perhaps even physiology, have conditioned us as subjective beings in such a way as to render thought about the hard problem persistently challenging. The mental is a map; the physical is a territory. There may be definable relationships between them (e.g., neural correlates of consciousness), yet we do not categorize them as the same *sort* of thing, in spite of the fact that the phenomenal world naturally appears to us *as* the external world. We *see* a territory and are poorly equipped to consider that what we see as but a map for the body’s use in navigating the world. *Facts* about the phenomenal realm can be mapped to *facts* about the physical realm, so that there can be some mapping between psychology and physics as domains of discourse; yet the phenomenal realm itself is not the physical realm itself. Nevertheless, the nature of their identification has long been a topic of debate within the philosophical community.[22]

The tension between “folk” psychological belief and scientific belief is built into the hard problem—for, philosophers and scientists are also folk as well as being professionals. There is a natural resistance to the idea that squishy brain matter can be responsible for consciousness, while it is easier (though probably wrong) to imagine that

[22] Cf. Heidelberger (2003, p255): “In his first publication on the mind-body problem, in 1934, after the general turn to physicalism, Feigl held the relationship between the physical and the mental for a *logical* identity between two descriptions of the given, a description in psychological and a description in physical vernacular… [In 1958] Feigl no longer saw the identity of the mental and the physical as a necessary, but as an empirical identity… The advantage of this theory is that it ‘removes the duality of two sets of correlated events and replaces it with . . . two ways of knowing the *same* event – one direct, the other indirect’.”

ants are conscious simply because we see them behaving purposefully (Arico et al, 2011, p349). Moreover, dualism remains a popular assumption even in the age of science, with many Americans and Europeans still claiming to believe in an afterlife (Riekki et al, 2013).

The gulf between physical and mental could even have a neurological basis. There is evidence of two separate functions, one for predicting behavior of inanimate systems and one for the moral sentiment or empathy involved in dealing with sentient creatures (Wright, 2007). Moreover, neither those biological mechanisms dedicated to perceiving the external environment, nor those dedicated to perceiving what is going on in ourselves and others, evolved to enable scientific investigation, but only to serve the needs of everyday living (Sloman, 2007, p10). Of course, having psychological and neurological roots does not diminish the hard problem, but it may help to explain why it is hard.

2.7 Philosophical contentiousness

A further reason for the refractoriness of the problem of consciousness is that much of philosophy remains concerned more with the analysis of language than with the analysis of nature. The hard problem straddles the border between philosophy and science, and has until recently been of interest more to philosophers than to scientists. Furthermore, many philosophers have been more interested in logical truth and the niceties of language than whatever physical truths lie behind consciousness.

But there is another sort of reason why many philosophical problems resist resolution. In the broadest terms, knowledge is inherently dialectical, since propositions can only ever be incomplete. Acts of analysis presuppose an existing unity and acts of synthesis presuppose a prior analysis (Talbott, 2004). The co-determination of experience allows theorists to emphasize either the object or the subject factor in their explanations, which spawns further divisiveness. Paradigm shifts in philosophy are not necessarily progressive, since philosophy is not an experimental science, with nature as arbiter to decide issues. Yet, even within science—and especially concerning consciousness—chronic irresolution of basic dilemmas gives rise to dialectical cycles in intellectual fashions and prevailing theories.[23]

Like other disciplines, philosophy is subject to specialization: saying more and more about less and less. Philosophers, who are professionally committed to argumentation just in order to stake out a niche, may be too contentious a lot to ever reach consensus regarding an explanation of consciousness. They love to take a position in regard to each other, often over rather contrived issues. One-sided viewpoints predominate; a mentality that insists on *either/or* cannot grasp the whole picture. For such reasons, philosophy often has an incestuous scholastic flavor, recounting who said what about who said what. It is more preoccupied with classifying concepts or theories, formalism, and nitpicking the relative positions of various expositors, than with

[23] For example, (Jack & Shallice, 2001, sec2): "A boom and bust cycle of consensus and controversy is evident throughout the history of scientific investigations of consciousness." See also (Velichkovsky & Pannasch ,2000, p1008): "Every forty years, somewhere in the world, there is a rebellion against passive, mirror-like theories of cognition."

understanding the real phenomena concerned. Splitting hairs creates entrenched camps. From tits beginnings in ancient Greece, philosophy has been a debating contest, prone to views based on a single aspect of a complex problem. The prospect in regard to explanations of consciousness therefore resembles the classic fable of the blind men examining the elephant.

Scientists enjoy continuing employment because the truths of nature may be inexhaustible. Philosophers enjoy continuing employment because philosophical argumentation is inexhaustible. There is progress in science because the mathematical-experimental program provides a common protocol and language, and because experiment (nature) can decide issues. There is no such arbiter in philosophy and no universal protocol or language except reason and convention. Perhaps in envy of hard science, philosophy tends to be dominated by a meretricious formalism and technicality. Some seemingly intractable problems in philosophy may persist simply because of the ivory-tower closure of the small community of professional philosophers, who define what the problems are and the proper terms in which to think about them and describe them. This may exclude new approaches and often seems to exclude the guiding force of common sense.

3.0 The context of the present approach

Two reasons for the renaissance of interest in consciousness involve technological developments: computation and brain imaging. The computer serves as a metaphor for the workings of the nervous system and also as a tool for modeling cognitive processes. Brain scanning technologies promise to identify neural correlates of consciousness.

The digital computer, as the universal machine originally conceived by Turing, is paradigmatically a disembodied information system, with text-based input and software, a machine that operates passively in a pre-determined manner. Even as such, it served to model real cognition as beginning with afferent nerves (input) and ending with efferent nerves (output), at the cost of ignoring the external environment (including the body) that is an essential part of the cycle of cognition (Shapiro, 2007, p340). On the other hand, a real computer is not actually self-contained nor a purely syntactic system. Even when it lacks an interface with the world through transducers and actuators (i.e., it is not a robot), its input is supplied, via users, from the real world and it serves some application in the real world, following human purposes. In contrast, a real cognitive system actively manages information for its *own* purposes, not those of human users. Perhaps nothing but further technological development prevents a self-organizing computer that does likewise.

The strategy to understand cognition in computational terms began with Descartes, the inventor of analytic geometry, whose early intuition was that propositional forms of representation (equations) can generate graphic forms of representation.[24] Early computer simulations of visual cognition involved interpreting visual scenes in terms of

---

[24] Descartes' great realization was that no likeness of the external world is transmitted within the nervous system, but only a symbolic representation (Huxley, 1874, p210). This is ironic given the association of his name with a "Cartesian theater" of inner images.

internal models, expressed as computer code, which provided a hopeful means to understand the brain's interpretive processes.[25] The computer is more equivalent to the brain in a vat, however, than to a brain serving an active living body. Moreover, the early approach to scene analysis was based on modeling cognition literally as a logical, language-based rational process. These shortcomings motivated approaches to embodied cognition that emphasize real-time interaction with a putative environment and de-emphasize internal representation. The idea is that the external world itself is its own best representation and store of information, which the organism can actively access for an update rather than consulting a laboriously constructed internal model. Alternatively, no external environment need be considered at all; the organism's task is simply to maintain certain relationships between input and output.

The embodied cognition paradigm acknowledges that an organism's agency cannot be accounted for by approaches that consider the nervous system to be self-contained, or to merely process information that is passively received. In accounts that lack the context of an embodied relationship to an environment, there is nothing 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. The enactive approach is a necessary corrective to the limitations of the older computational approach. But, it is no more sufficient to solve the hard problem than the traditional computational metaphor, because both are framed in extrinsic terms that fail to account for phenomenality (Ellis, 2000, p40; Rupert, 2015, Sec 6). Moreover, the controversy between representationalism and enactivism is another example of exclusive thinking: it is not a question of either, but of both—and more.

3.1 Representation versus enaction

The external world serves as an available reference for continually updating internal models, and the organism is clearly an embodied agent; but without some form of internal modeling to update, there could be no interpretative process to make sense of sensory input. Some may hold that there is no *need* for interpretation or modeling: the organism can survive just by managing input-output relations. This might be so for very simple creatures, but it is no basis on which to understand phenomenality.

Enactivism and cognitivism each emphasize a complementary aspect of cognition, but even combined they are unable to explain why it is "like something" to be in a phenomenal state. The reason for this is that phenomenality is a first-person experience, whereas the terms of both cognitivism and enactivism are third-personal. Describing events in the world (whether the external world or within the cognitive system that is part of that world), they fail to bridge the gap with the phenomenal experience to which those events correspond. There is still something missing in theories that remain, implicitly or explicitly, third-person accounts. Even considering interaction with environment, it is

[25] Cf. (Hoffman, 2011, sec1.2): "If the computer is going to see anything more than just this meaningless array of numbers, then it must have sophisticated programs that set about to construct visual worlds from the video input. Writing such programs has proved exceedingly difficult, and has led to great respect for the constructive powers of biological vision systems."

never quite clear how bustling neurons (let alone transistors) can produce the experienced greenness of verdant foliage or the feelings that imbue values, pains, and pleasures and which motivate actions. What is missing is some strategy to bridge the gulf from third- to first-person accounts. To explain the correspondence between brain states (or machine states) and experience, one must retrace the steps through which the cognitive system itself actively *makes* that correspondence.

Using the world as its own best representation seems reasonable as a path to action that requires no intervening consciousness (such as reflexes). It may work well to explain insect behavior (Dreyfus). However, beyond explaining behavior, the whole import and promise of internal representation is ultimately to explain *phenomenality*. Insects may not require internal representation (nor phenomenality), but humans certainly do. Sensorimotor accounts may justly criticize representational or computational theories of mind for considering the brain disembodied. But they throw the baby of internal modeling out with the bath water of disembodiment.

Just how valid the computational metaphor may or may not be depends on how "computation" is understood. Certainly, there have been great advances since "good old-fashioned AI"—such as "neural networks," "large language models," "predictive processing," and "integrated information theory." To reject *any* meaningful sense of internal modeling or representation was a move of desperation, leading to denial that the brain is responsible for experience at all![26] Rather, internal modeling should be integrated into a more comprehensive theory that includes referring to the environment as model (Grush 2003, p53).

The cognitivist tradition, based on the computational metaphor, first attempted to model cognition as a logical process of symbol manipulation. However, it is not necessary to take the organism's "logic" or "symbols" too literally in terms of human reasoning and communication—even when the organism concerned happens to be the brain of a human being. Similarly, "representation" can take many forms and should be understood as symbolic in a broad sense, and not as literally iconic, or as isomorphic to some external reality. Representation, abstractly understood, is a mathematical mapping between domains. These connections might be internal or between the brain and select aspects of the world (Crick & Koch, 1992/2002, p.11).

3.2 Kinds of representation

While the behavior of organisms without nervous systems may not involve internal representation (nor even sensory information, but only direct interaction with an environment), the mediation of some form of internal modeling seems necessary for the cognitive functioning of creatures with higher nervous systems, and certainly for conscious experience. While this does not imply an isomorphism of the representation with external reality, it does suggest agency and a mapping procedure, which may be regarded as internal communication.

The concept of 'representation' is no doubt derived from conscious human acts such as painting a representational picture or writing symbols to represent meanings or

[26] O'Regan & Noë (2001, p970): "The solution to the puzzle of understanding how consciousness arises in the brain is to realize that consciousness does not in fact arise in the brain!"

experiences. So, one usage of the term implies an intended surrogate for something else. This usage suggests that the representative image or symbol is an intermediary between the external world and the private realm of experience and thought. The representative painting implies a scene that is its real-world counterpart; but it is also an artifact with a real-world existence of its own. There is a tendency, then, to reify mental representation as some sort of *thing* inside the mind or head, leading to various troublesome corollaries such as sense data, homunculi, and Cartesian theaters. (Someone is needed to experience these inner representations.) But this defeats the purpose of representation in the first place, which is to explain how experiencing happens at all. Moreover, given that the brain is sealed inside the skull, 'representation' is a misnomer, since it is not a *copy* of the world but more like a *theoretical model* of it, which can only be formulated through interaction with what an outside observer takes to be an external world. Such a model entails the evolutionary history of such interactions.

In addition to the representing that goes on during conscious experience, there may be unconscious representation as well. (Consciousness implies representation, but not vice-versa.) Yet various usages of 'representation' are not clearly differentiated—particularly whether they refer to conscious thought and imagery or to abstract operations in a symbol manipulating system.[27] In the broadest sense, anything that functions to map another domain can be said to "represent"—in the abstract mathematical sense of mapping. Unfortunately, the term often carries more specific connotations, such as when it is construed to be a matter of recovering or reconstructing external features (Varela et al, p136). However, representation only has to involve operations that do not preclude evolutionary success.

Representations may be propositional or pictorial, digital or analog, and may employ spatial or temporal order and other relationships and patterns, so long as there is some set of elements with some ways to transform them (Sloman 2011, p18). The actual types used in a system—whether spatial (topographic) or functional (topological), for example—may be an empirical matter to discover (Pessoa et al 1998, p742), but with the caveat that while physical structure may *suggest* a particular logical structure to the observer's eye, logical structure remains always a matter of the organism's own specification. The system of representations may effectively constitute a language (Boden 1970, p205-6), but it is the organism's own internal language, not that of the observer. Any language or system of representations entrains its own assumed ontology (Sloman 1994 sec12).[28] It should never be assumed that this corresponds with the observer's ontology.

An iconic representation may be constructed from propositional information (as a monitor image is constructed of pixels), and this image may also then give rise to higher-level propositions, which may in turn give rise to higher-level iconic representations, and so on. There is thus a dialectical relationship between digital and analog representations. In no case should the iconic representation be confused with an *optical* image. An optical

[27] This reflects the trouble with the metaphorical nature of language, which is also its advantage.
[28] Cf. Sloman (1994, sec8): "E.g. predicate calculus presupposes an ontological carving up of the world into objects, properties, relations, functions, etc. plus a pair of truth values…" The "logic" of the organism as an information processing system is its own, and its "reasoning" is simply the "manipulation of information-bearing structures to obtain new structures that enable something to be done that could not be done directly with the originals…"

image falls upon the retina, for example, whose receptors effectively digitize it. The iconic image is a logical synthesis of the propositional (digitized) information.

3.3 Conceptual limits

The brilliant Leibniz could not conceive how the brain as a mechanism could give rise to subjective experience, and the challenge remains to this day. At fault may be our simplistic idealizations of matter, in contrast to romantic notions of our idealized being as subjects. The mechanical models familiar to Leibniz and Descartes were hopelessly simple. A modern computer is unfathomably more complex than a wind-up clock, yet is nowhere nearly as complex as a single-celled organism. It is possible to imagine that an artificial system does all the same information processing as an organism. Yet, we may not be able to imagine the complex processes whereby it does so *in sufficient detail* to properly equate their functional equivalence, let alone reproduce it. Part of the psychological problem may also be that we still intuitively conceive machines as material and minds as immaterial. An understanding of computers as *virtual* machines may help to overcome this obstacle. For, a virtual machine is a *logical* system even if implemented in a physical system. As such it is immaterial, and so may serve as a bridge across the explanatory gap. If one step closer, the task is then to show how a *virtual* machine can have experience and a point of view.

Now, the theory of the object offered by physics is incomplete without considering the role of the subject; and there is not yet an adequate theory of the subject.[29] The observer in physics is but an idealization, standing outside the system studied and exempt from the laws of nature. A complete science would have to take into account every aspect of the cycle of knowledge creation by including the observer. Our inability to explain the mental in physical terms is due partly to built-in limitations of physical explanation as presently conceived (Maxwell, 2000, p.50).

Our notions of the physical are derived from our creaturely experience on the mesoscopic scale. We extend this limited perspective through instrumentation and abstraction, which draw unevenly upon the various senses. The overall tendency has been to interpret data through the visual sense, in the wake of Descartes' initial program to reduce everything to extension in space (kinematics). Yet, dynamic concepts such as force and mass refer also to touch and proprioception and have remained troublesome for that reason. From the organism's point of view, what is "real" is what significantly impacts its surface and bulk, which the organism itself perceives.

The notion of "substance," as the intuitive essence of matter, has a tortuous history in physics. Attempts to reduce dynamics to kinematics aim to avoid the problems associated with it. At the quantum level, characteristics that render matter real disappear: individuality, locality, etc. Mass, supposedly the essence of substance, is interchangeable with energy, whose essence is motion. Even the measure of mass is defined to be independent of substance: behavior in a gravitational field, exchange of energy or momentum, etc.

---

[29] Donald Hoffman has proposed such a formal theory of perception, which takes the observer as fundamental, but at the price of resorting to thinly disguised idealism along similar lines to Leibniz' monadology. See: Bennett et al (1989).

Our notions of the mental are also derived from ordinary human experience, abstracted to concepts of soul, spirit, or ghost in religion, and to concepts of mind in philosophy and of the ideal observer in physics: a voyeur who can observe without being observed or influencing the system observed. Such an idealization works in the macroscopic domain as an effect of scale, where light hardly impacts the things observed (again, the paramount visual sense!) and where the transmission of information is virtually instantaneous. This circumstance breaks down, of course, in the microscopic realm, where the energies of the information transfer are of the same order as those of the things observed, and in the domain of extreme motion where the finite velocity of light cannot be ignored.

Our notions of complexity and simplicity are similarly relative to our creaturely scale and our natural mental proclivities. What is the ideal of parsimony but a psychological preference for simple expressions? There is no a priori reason to believe that nature itself is simple. It may appear to conform universally to simple laws—at least on the very large and very small scales—but this begs the question of whether such laws are formulated primarily *because* of the preference for simplicity that leads to idealization. Despite this preference, it appears that matter on our own scale can be extremely complex.

Overdetermination and multiple and circular causation may be the rule in nature; it is no mor than a metaphysical bias that there should be only single causes operating in a single direction. The quest for parsimony leads some researches to consider only the role of low-level phenomena that can be described in the idealized terms of physics (Sloman & Chrisley, 2005, sec 13). If the premise of physicalism is that mental phenomena can ultimately be reduced to deterministic physical terms, we should bear in mind that even many complex physical phenomena cannot be reduced to such terms (Ney, 2016).

Reductionism can nevertheless appear to succeed where structure and function are concerned, since these are paradigmatic objects of third-personal description and analysis. They are also primarily visual concepts. Information inferred or abstracted from phenomenal experience can be expressed propositionally, but qualia in themselves seem to resist such analysis. The problem is to reduce qualia (of whatever modality) to structural and functional terms; but this means inherently to terms of the visual modality, which becomes paradoxical when dealing with visual qualia themselves. In fact, as we shall see, qualia *do* bear structural and dynamic information—about the external world, to be sure, but also about the organism, its sensory organization, intentions, values, evolutionary history, etc. But this latter type of information is expressly factored out of what is typically considered essential in physics, whose conceptual framework is essentially formal and mathematical.[30] At least in terms of their idealizations, we can analyze both ordinary and scientific objects in structural and formal terms, to the extent we can physically and conceptually take apart such entities. But qualia resist such analysis because they are gestalts in the first place: first-person integrations or syntheses of structural information. Furthermore, we have no conscious access to those structures

---

[30] Cf. Lockwood (1998, p88): "The overtly algebraic character of modern physics makes this situation plain. It is clearly impossible to get an illuminating answer to the question of what, for example, the variables in Maxwell's equations *really* stand for. If one grasps the mathematics and uses it to generate predictions about what one will experience in various situations, that is all the understanding one can reasonably hope for."

and processes that lead to the integration, parallel to the conscious access we have to scientific methodology. We cannot conceptually take apart that of which we are not even aware.

How phenomenality "emerges" from neurological events or states is sometimes likened to how liquidity emerges from molecular motions and states.[31] However, liquidity is an ambiguous notion; it is a phenomenal quality as well as a phase of matter. As phenomenal quality, liquidity consists in such experiences as the tactile feel of wetness and viscosity, the visual appearance of fluidity, the sound of flowing or dripping, etc. As such, it presents the same challenge as color experience—namely, to explain its emergence from *neural activity*, not from microphysical properties of the matter in question. For, we can perhaps explain the *difference* between the appearance of solidity and the appearance of fluidity, but not the appearances themselves. Similarly, it may be possible on physical grounds to explain why a certain neural pattern produces a certain experience *rather* than another, but not why it produces any experience at all. As Chalmers (1995, p20) puts it: "A physical theory gives a theory of physical processes, and a psychophysical theory tells us how those processes give rise to experience. We know that experience depends on physical processes, but we also know that this dependence cannot be derived from physical laws alone." In those terms, to solve the hard problem requires a psychophysical theory rather than a purely reductive physical theory, such as how liquidity arises from the interactions of molecules.

We have a natural tendency to imagine physical reality on any scale in the ways familiar on our mesoscopic scale. In the mind's eye, at least, we picture molecules and galaxies much as we picture chairs and boulders. Yet, we know that such appearances, of things on the largest and smallest scales, are but acts of visual imagination, *as though* from the actual perspective of a naked eye.[32] We know also that we have categorized things in such a way that what we think of as *qualities* on the mesoscopic scale, such as wetness and color, are supposed to "emerge" from what we think of as *properties* on the microscopic scale. In the scientific program, reduction to extensional or quantitative terms leaves us with the need for some procedure for the mind to reconstitute the qualitative facts of ordinary experience from theoretical facts of science. That procedure necessarily implies a subject to perform it, who has the ordinary experience in the first place, from which the theoretical facts of science are abstracted. The *objective* phenomenon of emergence—of macroscopic properties from microscopic properties, such as liquidity from molecules—can ultimately not be dissociated from their emergence as qualities (such as the feel of wetness) in the consciousness of a subject. 'Emergence' can fill a legitimate explanatory role when novel properties on one scale can actually be derived from known properties on another scale. But, it may also simply be

---

[31] However, absolutely pure (theoretical) H_2O is an idealization only approximated by normal water, whose bulk properties cannot be predicted from the molecular properties of theoretical H_2O alone (Beaton, 2016, p364).

[32] While some planets and galaxies present extended visible images, which can be technically enhanced, others are too distant for present technology to treat as more than detection events. Similarly, at the limits of microscopy, the difference between image and detection event is relative.

used as a fudge factor to name a gap when this derivation is not feasible. In short, it is a relative and subjective term (van Gulick, p20; Revonsuo, p26).

The theoretical constructs of physics, based on experiment and observation, are no more independent of the minds of physicists than folk categories, based on everyday experience, are independent of the minds of ordinary people. One might prefer to think that the entities of the scientific world are what really exist and that the phenomenal world is a kind of biologically useful illusion. On the other hand, one could justifiably hold that the manifest world given in experience is what is real and that the entities of science are but pragmatic artifacts, symbolic tools to help us survive (Sellars, 1963, p31). A third possibility is see them both on the same footing: the image of the world given in normal perception *and* the scientific image of the world are alike useful constructions, virtual representations. If there seem to be *only* appearances of one sort or another to deal with, that is only to say that the concept of reality for a self-conscious being is unavoidably paradoxical. In particular, theoretical constructs (including the very idea of an objective external world) enable us to function in the world, just as our everyday perceptions do. The hard problem is grounded in tacit assumptions about the nature of physical reality to which the phenomenal realm is supposed to be reducible. It may appear less hard when we count those as mere assumptions.

3.4 Virtual machines

While the brain is undeniably a physical system, its mental aspect derives from further being a control system. To the extent it can be considered in mechanistic terms, the brain is a *virtual* machine. The computational metaphor serves well in this regard at least: the brain, like the computer, must be understood in terms of logical rather than causal connections. Once that is understood, there should be no question of reducing one category to another. The way is open to a different kind of analysis, in which intentional phenomena on one level can be assimilated to intentional phenomena on another level, with no categorical gulf to bridge. We need not be puzzled that a virtual machine can also be physical.

The components of an information processing system are virtual components, even if they happen also to be physical. Observing the behavior of such a system is a different task from observing the behavior of a strictly physical system (Sloman & Chrisley, 2005, sec1). It requires a different stance (Dennett, 1987/1998). The neural correlates of consciousness are analogous to the hardware correlates of a virtual machine. Physical structure may *suggest* logical structure (to the observer's eye), and vice versa; but, where organisms are concerned, the logical structure is specified by the system itself. One may speculate freely about which virtual machines can best model the human brain. Whatever the answer to that question, the processes in a virtual machine can legitimately be said to cause physical effects, so that downward causation is feasible and epiphenomenalism is false (Boden, 2006, p1421; Sloman, 2004, sec4).

The notion that consciousness is a natural kind of virtual reality, an internal simulation of the external world, is a logical extension of the computational metaphor (which in turn is a refinement of the mechanist metaphor). It is a fruitful heuristic, but its limitations as a metaphor must be kept in mind. Literal VR is an *external* input, like any other sensory input, enjoyed by people with real brains and bodies already engaged in perceiving external reality. To explain the perception of external reality as an internally generated simulation (which includes the presence of a virtual self) unavoidably involves circularity (Westerhoff, 2016). Some authors take the metaphor to an extreme, likening waking life to "online dreaming" (Metzinger, 2005, p7). Here we use the concept of virtual machine not to argue for idealism, or that phenomenal experience is illusory. We simply wish to demonstrate that an understanding of consciousness does not require, and cannot be achieved by, reducing mental phenomena to strictly physical terms.

4.0 The challenge for the theorist and the challenge for the organism

From the point of view of the observer, the organism is a black box that strives to maintain a set of input/output relations. From the point of view of the organism, the world is a black box to manipulate in such a way as to maintain those relations. An observer may view the organism as adapting either to an environment or to changes in its own states. However (or, indeed, whether) the organism perceives its environment, the observer is tempted to view its actions as taking place upon what the observer considers that environment to be. While it is understandable to identify and interpret the structure and functioning of organisms according to human categories, definitions, and purposes, the very nature of the organism is to be *self*-defining, to have its own priorities. It is only incidentally an object of human definition and study. The challenge is to understand the logic of the organism's cognitive self-programming—*from its own point of view*. Indeed, the challenge of the hard problem is to understand how it creates that point of view.

The notion of the operational closure of a system (modeled on the mathematical concept of closure) is that "the results of its processes are those processes themselves" (Varela et al, p139).[33] From the point of view of the observer of an organism, the cognitive loop which thus closes upon itself includes an environment from which input is drawn and upon which output is executed. The autonomy of the organism, defined in terms of this closure, does not mean independence or disconnection from an environment (solipsism). The organism itself may view its environment differently (or not at all, if it simply responds without awareness to changes in what the observer takes to be an environment). Some enactivists make the point that this closure obviates in principle the need of the organism to consider—and even the possibility to *represent*—an external world. Instead, a cognitive system will "*enact* a world as a domain of distinctions that is inseparable from the structure embodied by the cognitive system" (Varela et al, p140). Such statements simply muddy the waters.[34] For, what else is a "world as a domain of distinctions" than a representation? And where is this structure embodied if not in the real physical world? Certainly, the organism makes its *own* domain of distinctions, which will reflect its own structure as well as that of what the observer takes to be the external world. But the fact that distinctions are made by and serve the organism hardly implies

[33] Strictly speaking, mathematical closure means that the results of processes are also processes, not the identically same processes.
[34] The term 'enact' suggests a fictional stage drama. As such it seems calculated to avoid reference to a real world (i.e., to avoid representation).

that they do not refer, deliberately or incidentally, to an external reality.[35]

4.1 Taking the organism's point of view

Just as realism concerning the external world cannot be naively interpreted as mind-independence, so autonomy of the organism does not mean isolation from its environment. But how to define what constitutes that environment? From the point of view of the observer, the organism selects the patterns in its sensory input that are significant to it, which become for it effectively its environment. The organism also helps to shape that environment and the world perceived by the observer, which may overlap in some ways and not in others. Organism and environment mutually adapt. Yet, the world normally perceived by a human observer should not be presumed as the environment for the organism. Nor can the observer presume that the environment for the organism is merely a subset of the observer's own perceived environment; for many organisms have sensitivities not shared by human beings.[36]

The observer defines the organism as a system distinguished from its environment by a membrane, while thermodynamically open to it and causally continuous with it. Indeed, (from the observer's point of view) the organism itself behaves in such a way as to set itself apart from, or in opposition to, an environment. While this is surely a biological basis of the subject-object split for human beings, who find themselves both observer and organism, it is also the basis upon which the organism can be regarded as an intentional system, an agent in its own right. As such it has its own purposes, on the grounds of which it makes internal connections, which are not necessarily those an observer would make.

The intentional and design stances (Dennett, 1987/1998) unavoidably indulge rationales of the observer. But the organism has its own rationales. One must take care to differentiate between human categories, reasons, and purposes and those that might be the organism's own. We are looking over the creature's shoulder, so to speak, as we would over a programmer's shoulder in trying to grasp the logic of the program. The difference is that one cannot take for granted that the "logic" of the self-programming organism bears any resemblance to that of a human programmer. It might appear that an observer can determine how or whether a system is internally representing an object by

---

[35] "Prediction Error Minimization" theory similarly emphasizes the autonomy of the organism in such a way that no *mention* of the environment need be made. In that theory, perception is a matter of the brain comparing internally generated models of the world with sensory input and acting in such a way as to minimize the difference. That is, the difference between a sensory input predicted by the model and the actual input is kept to a minimum through constant feedback, which implicitly results from action upon the unmentioned environment. Cf. Sterpetti, Sec 1.

[36] For instance, birds can see beyond the human range, into the ultraviolet, which enables them to identify each other though they look the same to humans. The fact that we can know of sense modalities beyond our own (e.g., sensitivity to ultraviolet wavelengths, polarization, or electric or magnetic fields) includes them our ontology, making it possible to have technological versions of them. It might seem therefore that, at the top of the food chain, our view of what constitutes *the* environment forms an absolute standard for other creatures. However, we cannot dismiss the theoretical possibility of advanced aliens with a more comprehensive concept of what we consider the environment.

observing its reactions to that object—that is, by interpreting the meaning of its behavior (Dretske, 1994). But meaning to whom? Ideally, we would like to know what it means to the system itself; however, we have only the tool of the intentional stance with which to second-guess that meaning—effectively, to imagine it in human terms. (This applies even to self-knowledge, when "one" supposes some subsystem or sub-personal agent within oneself which is doing the representing.) Mental terms such as 'reasoning', 'decision-making', 'thinking', and 'representing' should be used with caution, since they refer by default to human conscious processes. Similarly, when considering evolutionary arguments, an outside observer evaluates utility implicitly in human terms. Nevertheless, the organism's perception must reflect its own estimation of utility, however that is to be framed. Though a product of evolution, the function of an organ or behavior is typically evaluated in the terms of the observer's ontology.

4.2 Formal systems, propositions, facts

Attempts to understand the logic and reasons of the organism defer by habit to human logic and reasons. (This applies even to understanding the bizarre reasoning of the human subconscious.) The significance of the computer metaphor for understanding the brain lies in applying conscious human rationality to model brain processes. Computer programs are based on formal logic, which extends and abstracts natural language—which in turn reflects and shapes the human thought processes it attempts to express. The advantage of formalization for understanding the minds of other creatures is precisely its abstractness, which (supposedly) frees "logic" from the idiosyncrasies of human thought and particular languages (Ayer, p88). This should enable us to entertain the idea that the organism may have its own characteristic reasoning processes that do not necessarily correspond to human rational thought. However, the tacit assumption may be that formal logic represents an ideal to which other minds can only approximate, just as the human cognitive domain and ontology may seem to set the ultimate standard for what can be perceived as reality.

The *proposition* is the formal version of the natural-language sentence, and the *formal language* or *formal system* is the formalized equivalent of a natural language (with elements that correspond to nouns, verbs, syntax, etc.) or a natural system (with elements such as objects, forces, and natural laws). To analyze an organism as an information-processing system is to treat it as a formal symbol-manipulating system and to translate its putative operations into propositions of a formal logic—equivalently, a program. This is perhaps the best we can do to free ourselves from human parochialism, with the caveat that the elements of such a formal system need bear little resemblance to human thought, so long as they satisfy the formal constraints, nor to what the human observer considers the real environment. In other words, we must take care not to project human reasoning into the intentional processes of the organism.

Trivially, human thought has propositional content—in contrast to human qualia, which supposedly do not. Yet, even qualia may be said to have such content implicitly, if it can be translated into propositional form. Conversely, the ability to derive propositions (facts) from a natural phenomenon does not mean those facts are true, much less that they correspond to the actual structure and dynamics of the phenomenon. Perhaps the notion of ‘propositional content’ would be better considered a way of looking: call it the “propositional stance.”

The extraction of facts (propositions) from phenomenal experience is what gives us a communicable notion of truth. But this extraction comes at the price that the experience itself cannot be reconstituted from such facts. (This is one way to state the hard problem!) To know what it is *like* to taste a pineapple is not the same as knowing *properties* of the pineapple—or even properties of the whole system including the subject’s physiology and brain. In truth, however, not only is propositional information derived from qualia, but qualia themselves can be said to express propositional information *in an integrated format*. The process of that integration is unconscious, so we have no conscious access to the information that the brain uses to constitute qualia. This renders the relation between thought and phenomenality chronically problematic, since thought (even mental imagery) seems a pale reflection of sensation (Talbott, 2004). Just as formal descriptions must be “interpreted”[37] to make experiential sense (and as words must evoke associations in order to have meaning), the task of the hard problem is to show how the brain interprets its own internal communication. But this is with the caveat that such internal communication can only heuristically be considered a language consisting of propositions (Dennett 1987/1998, p206-7).

From a formal point of view, intentionality (as we shall define it) involves symbolic operations that in themselves are gratuitous, purely syntactic. An observer can understand the cognition of an organism as a potentially formalizable, free-standing, information-processing system. The observer, and indeed the organism itself, may then interpret this system to refer to the real world. But their interpretations will almost certainly not be the same. As a sheer formalism, a cognitive system is abstract, self-contained, and arbitrary. (The brain that supports it is accordingly often viewed as isolated and passive, like a computer.) Yet, like language and mathematics, a real cognitive system can refer outside itself to create meaning for the organism. A formal system must be interpreted in real-world terms to have meaning or be of use, just as a language normally conveys meaning through semantic reference. Just so, embodied cognitive systems come to refer outside themselves through interaction with an environment. Such reference is an act of the organism, in the context of *its* world, not that of an observer. Hence, to arrive at a program that models mental functioning, one must consider how the organism programs itself, whether that means at the level of the individual creature or that of the species. 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-programming—through the organism’s evolutionary history, through the ontogenetic history of the individual, and through its ongoing interaction with the world as a learning process.

4.3 Models and autopoietic systems

If the organism has sense organs, it responds to changes in these through activity that restores a preferred state. Even without dedicated sense organs, it may respond to

---

[37] In the mathematical sense of “interpreting” a formal system. Geometry, for example can be a purely abstract formalism, which can also be “interpreted” in terms of physical space.

changes of its own chemistry, for example, in ways that prove adaptive or not. Either way, the organism is effectively "flying by instrument" (Oatley, 1978; Maturana, 1978). The single cell and the human brain sealed within the cranium face the same challenge: to respond to disturbances in ways that restore its equilibrium and favor its continued existence. From an observer's point of view, there is a world external to the cell wall, the skin, or the skull. However, the challenge for the organism is survival and reproductive success, which may or may not entail activities that the observer interprets as modeling an external world.

The computer metaphor is a tool for understanding cognition from an engineer's, point of view. It involves applying a rational design stance to processes that might have a completely different rationale than conscious human thought. The challenge of modeling nature is not limited to modeling living or cognitive systems, of course, but is more problematic where the system concerned can be considered an agent and not merely a passive transmitter of efficient causes. The mechanist worldview allows the observer to specify input to a fixed system, producing a determined output. However, the organism defines its own input and the rules of its processing system, which may not correspond to the input and system as defined by the observer. The organism also changes itself in the process, so its own organization is part of its output.

It is possible to conceive a system that does "all the same" information processing as humans (philosophers seem fond of doing this). But it is also possible that we cannot imagine it in sufficient detail, or in the right ways, to correctly recover the functionality of those processes associated with phenomenal experience or with some sophisticated behaviors. Simply to identify, or even to duplicate, structure is not enough to be convinced that the processes are truly identical or even analogous. After all, the computer itself demonstrates that the same hardware can function in vastly many different ways! Since the brain is not an artifact (nor literally a digital computer), it may not be possible to model its structure with confidence. For, scientific models are inherently finite, simplistic and well-defined, while nature is inherently complex and indefinite. Any functional or structural analysis of a system (a model) is itself an artifact, equivalent to a program or machine, with a definite information content. It cannot be proven to be isomorphic to any natural reality, which has no definite information content because it cannot be uniquely defined.

We cannot understand the behavior of the organism (including perceptual behavior) without knowing something of its goals, values, and strategies—its "software," so to speak. The reality of these is as elusive as its putative hardware. The organism has its own intentionality, which is not the same as ours. A scientific model is a product of human definitions, while an organism self-defines. Its own internal modeling (if any) will not necessarily correspond to the scientist's modeling of its modeling. We can acknowledge the differences between our terms of reference and the organism's.

A vexing problem for neuroscientists and AI workers, who wish to simulate the nervous system (let alone the whole organism) by means of digital computation, is that an adequate simulation may require formally listing an indefinite number of propositions. Despite the speed of computers, this can lead to massive computation with unrealistic slowness compared to real-time interactions of creatures with their environments. Overcoming this obstacle is one motive behind other approaches to computation—and also other approaches to cognition, in which behavior is linked more directly to input from the environment, avoiding the computational bottleneck implied in internally stored representations. Somehow the real brain overcomes such limitations. Short of throwing out the notions of representation and computation entirely, we must admit that our human, proposition-based notions of computation and representation may be poor guesses at how it does this.

In many ways, nature seems committed to efficient ways of doing things. In the case of organisms, another way to say this is that the living creature has its own rationales and methods that obviously work. These may no more correspond to propositional analysis (and, hence, to computation) than our own unconscious processes correspond to conscious thinking, which is vastly slower. The organism does not need to do things *our* way, let alone perceive reality "objectively," in order to survive. Nature appears to be an intelligent designer because only viable designs have survived the culling of natural selection. But even that is relative, since most designs that have ever existed worked only for a time, in a given niche. None of nature's designs come with a blueprint or user's manual to reveal the true functions upon which functionalism would ideally rest its case (Dennett, 1987/1998, p321). This does not preclude a function, meaning, or intention established by the organism itself (or by "Mother Nature"), whether or not we can know conclusively what that is.

5.0 Traditional intentionality

The literature on intentionality has mostly dwelt on the notion as Brentano considered it—as a function of language. But it is language that should (ultimately) be explained in intentional terms, not the other way around. Moreover, as developed by phenomenologists, intentionality regards only the conscious human subject. Since it is consciousness that we hope to explain, the notion of intentionality must be broadened.

The language-based concept of intentionality developed by Brentano is primarily about reference: the signification (to human beings) of signs. The significance does not inhere in the sign, however, but is a mental act. We cannot deal with *objects* of reference without mentioning an agent who does the referring. If intentionality entails an object, it must also must entail a subject: an agent who enters the particular relationship to the object involved in cognition. Within that framework, qualia are as "intentional" as propositions, beliefs, and desires. There is something they refer to or are about; they convey information.

Wikipedia defines intentionality as "the ability of the mind to form representations," and warns that it should not be confused with intention in the ordinary sense. I propose to regard intentionality as the ability to form "intentional connections," a notion not to be specially identified with human language, propositional thought, or external reference. The brain (or organism) is an intentional system in this fundamental sense of being an agent that makes internal connections for its own purposes, regardless of whether these refer to anything or constitute representations, and regardless of what form representation may take.

Wikipedia continues: "Thus, a thought of a chair can be about a chair without any implication of an intention or even a belief relating to the chair. For philosophers of language, what is meant by intentionality is largely an issue of how symbols can bear meaning." Fair enough. But "symbols" do not have to be elements of human languages, which are features of the observer's ontology. Likewise, "belief" and "intention" are normally concepts in the specifically human cognitive domain. To apply to mind in a more general way, and to include other creatures, I prefer to substitute the more general notion of *intentional connection*, which is essentially a mapping, in the mathematical sense. Insofar as an agent has its own reasons to make that connection, the thought of a chair *does* reflect an intention, perhaps even a belief.

5.1 A different notion of intentionality

The transitive Latin verb *intendere* can mean either *to mean* or *to intend*. Brentano's intentionality takes off from the former, whereas common English understanding of intention refers to an action and its motives rather than just to words, thoughts, and meanings. I propose to follow this sense of action, which necessarily involves an agent making a connection for a reason. Even if the action is mental, logical, or virtual, the focus is on the agent as well as the recipient of the action, which is not limited to symbols. To intend, in this very general sense, is simply to *make* such a connection, whether between a present state and a future one, between a sign and what it signifies, or between one existing connection and another. The connection may be entirely internal. Such an agent does not have to be conscious, nor does the intentional connection have to be consciously made. (Indeed, it should serve as the basis to *explain* consciousness.) *Agency* in this sense is a fundamental concept, not limited to conscious human beings and their molar actions. An intentional connection is thus fundamentally different from a causal connection, though they may describe the same event.[38]

The traditional notion of intentionality as directedness or "aboutness" is often understood in terms of grammatical language, with transitive verbs and their objects. A broader understanding of intentionality is here proposed, based not on the linguistic relationship of *referring* but on a relationship ordinarily known as *intending*. By this definition, *all* mental processes are intentional, insofar as they involve connections made in a space of reasons. The emphasis is shifted from external reference to internal connection, whether or not consciously made, though it may be interpreted by outside observers as referring to the external world. Intentionality, in this sense, can perhaps best be understood as mathematical mapping or transformation, from one domain to another. It thus accommodates the "directedness" of language without being tied to language use.

The connection between elements of a symbol system and whatever they symbolize is *intentional* (in the commonly understood sense of serving a purpose) in contrast to being *causal*.[39] The symbols in themselves are arbitrary; their connection to

---

[38] An agent makes an intentional connection within itself, or between some element within itself and some element in the world. An observer (who is also an agent) makes an internal connection that represents a connection that is supposed to exist between elements in the world, and calls it a causal connection.

[39] Brentano's "intentional inexistence" or Chisholm's "power to misrepresent", for example, effectively mean that a symbol system is *not* a causal system, in which one element is bound by some natural necessity to another. On the present theory, elements can be arbitrary, bound instead by the action of an agent who makes the connection between them.

anything else is not established by any natural or metaphysical necessity, nor by anything inhering in the symbols themselves, but by an *agent* acting with its own purposes.[40]

Intentionality does not by itself imply phenomenology, since it underlies both experience and behavior, both conscious and unconscious processes. Thinking, believing, perceiving, desiring, intending, deciding, inferring, etc., can all be interpreted behaviorally without invoking phenomenal experience. Yet, there clearly is a phenomenology of many mental activities normally considered intentional in Brentano's sense. There is actually something it is like to be willing an action, thinking a thought, believing or intending something, etc. While intentionality has traditionally been associated with propositions and not qualia, I have argued that qualia *do* bear (propositional) information about the external world, as also recognized by the notion of "phenomenal intentionality" (Lycan 2008).

5.2 Agent *vs*. patient

The notion of *agent* complements the archaic notion of *patient*, though this is no longer a distinction made in natural science, which has defined matter in principle to be essentially passive and physical processes to involve only efficient cause, never the teleology of agents.[41] Agents serve as *initiating* causes (if not ultimately *first* causes)—in contrast to *efficient* causes, which passively transmit dynamical processes through a system from a source outside, and which involve energy exchange, force, motion, etc., though not intention or teleology. (In Aristotle's terminology, the goals and reasons of agents are *final* causes, and *do* involve teleology.) Of course, to exist in a real system, intentional connections or processes must coincide with physical connections or processes. Nevertheless, an intentional connection is a *logical* state as distinguished from a *physical* state; it is made by an agent using its own initiative and energy resources for its own purposes. The transmission of causal processes in the world becomes in the organism a transmission of "differences that make a difference," powered by the organism's metabolism (Bateson, 1972, p453).

Computer are designed so that their states can be made to represent propositions. The brain can be understood functionally to do something similar, though its "commands" and "propositions" are not to be understood literally as human language. In the case of the computer, the commands or propositions are not its own, but the programmer's; they have no meaning to the computer itself (to which they make no difference). The computer is a human tool, not an agent in its own right. In contrast, the organism is an agent whose processes serve its own purposes. Its internal operations can be viewed as propositions or commands, on the model of computation, where the

[40] Cf. Kuipers (2007, p86): "The ability of a symbol to refer to a distinctive state in the physical environment depends on the behaviors of the… [agent], not on any pre-existing intentionality in the set of symbols."

[41] Psychological experiments support the notion of a basic human category of "agent" or "intentional agent" as a sufficient basis for attributing phenomenal states (Arico et al, 2011, p332).

organism itself is the programmer. Such operations should be understood as internal connections and directives made by the system itself, not by an external programmer.

There is a categorical difference between intentional connections and physical connections, since 'intention' and 'cause' are distinct relational primitives (Mausfeld 2002, sec4). Physical cause is generally understood as something that "happens to" a system, or happens within it but is initiated from outside it, and which changes the state of the system (the effect). The cause of this change may be an agent outside the system (in some cases, literally an experimenter). But normally cause is considered an event in a larger system that is also merely undergoing passive changes. The system's borders can be extended indefinitely to the boundary of the universe, so that there are no agents within the defined system, only efficient causes. But this is at the price of a final reckoning in which agency is deferred to a first cause outside the universe. Dealing exclusively with closed systems and efficient causes seems to avoid the dualism of matter and mind, but the eventual necessity of a first cause to set the chain of efficient causes into motion simply pushes the mental outside the boundary of the system studied. (For the early mechanists the first cause conveniently was God.) When that system is the universe as a whole, there is no place within it for the observer to stand and no natural place for a first cause to act from. While this inconsistency is normally ignored, one way to face up to it would be to review the privileged role of efficient cause within scientific description.[42] The understanding of matter as inherently passive suited the study of inorganic systems. It fostered the development of technology, of physics as the paradigm science, and the exploitation of "natural resources." It never did suit the study of life or mind.

An obvious objection to admitting agency and teleology into scientific discourse is that one might not know where to draw the line, or on what basis. (Consider the current confusion over whether Large Language Models should be treated as minds.) It was partly to avoid this eventuality that science was defined as mechanistic in the first place. However, neither machines nor the inorganic world in general are now considered to be fundamentally as simple as classical mechanics held them to be. As the distinction between organic and inorganic becomes progressively blurred, it will be necessary to allow agency within nonliving matter. Self-organization is a concept that becomes relevant in ever more domains. While it cannot fully be developed without the related concept of agency (Bruiger, 2017), the task will be precisely to know where and why to the draw the line: on the basis of autopoiesis.

The brain is at once a causal and an intentional system, because it is part of an autopoietic system, an organism. Whereas there has seemed to be a tension between the self as physical and as spiritual or mental—and a category gulf between mind and matter—there lies no contradiction in the organism being both agent and patient, at once embodying two notions of causality. Nothing in principle prevents intentional and causal descriptions from referring to the same physical system.

On the other hand, excluding the subject or agent from discussions of intentionality leads to unnecessary confusion. What accounts for the directedness of thought, for example, is not some inherent "directable" property of objects of thought but the agent who directs. An intentional system *acts*, for reasons which are final causes,

---

[42] Efficient cause was but one of four types of cause considered by Aristotle; the others were rejected for historical reasons by the founding fathers of the scientific revolution.

while a causal system *reacts* to efficient causes. The intentional system acts in its own domain of reasons, and the essence of "mind" is "action in a space of reasons" (Beaton, 2009, sec2.3.1). Such reasons are the system's own, not those of the human observer. On the other hand, it is the observer who makes attributions of intentionality or causality.

How, then, should we understand the relationship between intentionality and causality? Here we propose that it is like the relationship between computer software and hardware. The physical system does not *cause* the operation of the intentional system, but underwrites it. Computer hardware does not cause software, nor vice-versa; rather, both supervene on the intentions of human designers (and beyond that, the needs of users and the marketplace, etc.) Similarly, mind and brain do not cause each other, but are caused in common by evolutionary factors.

To explain the arising of phenomenal experience within a physical system, we must look to its actions within its space of reasons, which means looking from its own point of view. While it might seem feasible to explain the system's *behavior* purely in terms of efficient causes, to ignore the system's own considerations would severely limit that understanding of behavior. It is not possible to understand *phenomenality* on such a basis.

The very notion of behavior is ambiguous, and is inappropriately restricted when it entails only physical description in the context of organisms. The behavior of creatures is more than physical motion: it is the action of an agent (Long, 2010). When causal description of a creature's behavior is viewed merely as a sequence of physical events in space and time, without regard to the creature's agency or intention, its behavior then may appear to arise in a deterministic way, proceeding inexorably from sensory stimulus, through electrochemical connections within the organism, to motor response. Such a molecular description can hardly account for the actions of an organism as an autonomous agent. While very such agent, being physical, is also a causal system, not every physical system is an agent or mind.

The notion of free will seems challenged by the systematic delay between neural events and a corresponding subjective experience of willing voluntary movement (as in the experiments of Libet). But, intentional connection takes place in *logical* order, not necessarily in temporal order: the if/then of syllogism rather than cause/effect in time and space. Of course, intentional connections must be embodied in a physical system to exist in the physical world. What counts is their *correlation* with physical connections or events, not the temporal order involved.

If intention can cause action, what causes the intention? This begs the question of what kind of causality we are talking about. For, the notion of efficient causation tacitly presupposes a first cause outside the system, whose initiating action remains itself uncaused, but whose effects are then deterministically transferred throughout the system like falling dominoes. However, the paradigm for an initiating original cause is the body itself, or the early experience of willing it to action. The very notion of physical cause is first acquired through an experience of "force," perceived by the body tactually and kinesthetically and initiated by voluntary motor action. However, this bodily notion of cause is soon generalized to the effects of external things upon each other, which can be inferred by *visual* observation. Hume objects that such observation reduces simply to sequence in time; whereas the notion of causal necessity (force) behind such sequences derives from bodily experience.

The intentional system makes intentional connections leading to action; the *observer* makes causal attributions, after the fact. An organism is part of a network of efficient causes perceived by the observer as well as being an agent in its own right. The flow of events in that network can be described either in physical or intentional terms. Similarly, knowledge of events can be described as a flow of information transmitted through intentional connections; but "information" can also be described as an extension of the events themselves, causally transmitted within the organism and across its boundaries.[43]

On one view, functionalism describes the organism and the world alike in informational terms (Boden 2006, p1434). The organism uses information, and mind may be likened to information processing; but information so conceived is essentially propositional, a third-person concept. To know facts (propositions) *about* pain, for example, is not to know the pain itself, because the information that the nervous system uses to *produce* the pain is not the information an observer has about it. The "information" involved exists in different domains, serving different agents. On one level, the cognitive agent passively suffers pain; on another level, it actively produces it as an internal communication. On a third level, another party (even, perhaps, oneself) simply observes the signs of it.

For several decades there has been a program to "naturalize" intentionality by mapping relations between brain states and states of the external world [Kriegel 2012, abstract]. The "natural" relation sought is by default efficient causation; for, a cause can be understood as the occurrence that provides maximum information (to an observer) about the effect (Sayre, 1976, p73). However, this is an observer's epistemic strategy; being empirical, it can only be probabilistic. The observer can only know what is "maximal" through some comparison after the fact. From the brain's point of view, its connections are simply true by definition, as acts of fiat.

5.3 Intentional systems

While signal transmission within the nervous system depends on causal connections (as discriminated by an external observer), the meaning to the organism itself rides upon connections made by, for, and within the organism. Neural connections are intentional connections in the sense I have defined. Whatever else it is, the nervous system is an *intentional system*, which I define as a system of intentional connections.

This differs from Daniel Dennett's (1978, p238) well-known definition of an intentional system as a system or object toward which one can successfully adopt the intentional stance for the purpose of explaining its behavior.[44] Dennett (1981, p221) thus

[43] There is a growing tendency in physical sciences to regard information as ontologically fundamental. Some therefore distinguish "intentional information" from "physical information" (e.g., Faichney, 2013, p42). The distinction parallels that between secondary and primary qualities. Physical information is structure, but structure is in the eye of the beholder. It cannot be discerned without some interaction with a physical observer. It is therefore questionable that physical information has an observer-independent existence.

[44] Dennett's *stances* (physical, design, intentional) are defined from an observer's (third-person) point of view. An intentional system, in his sense, is one whose behavior the observer can predict

makes the intentionality of the system a function of the observer's strategy and intentionality. I prefer to define the intentional system as one having its own (i.e., intrinsic or original) intentionality. The question of whether a system's intentionality is intrinsic or derived is not about a *kind* of intentionality (derived or original), but about *whose* intentionality it is—specifically, whether it is the system's own or the observer's.

I propose to take the notion of intentional system as fundamental. In such a system, intentional connections *define* the system while physical connections merely carry them as material cause (as Aristotle would say). For example, the intentional connections within an electrical artifact are specified by the logic of its wiring diagram, not by the physics of its soldered wires. The task is then to explain human rationality or intelligence, for example, in terms of intentional connections that are not necessarily "rational" in the terms of conscious human thought.

For Dennett, whether something is an intentional system is a matter of how one looks at it. Yet, one can say the same thing about causal systems: a particular thing is a causal system only in relation to the strategies of someone who is trying to explain and predict its behavior using the physical stance (i.e., in terms of efficient causes rather than reasons). But, as we have seen, even the chain of efficient causes ultimately presupposes an (intentional) agent outside the system, to serve as first cause.[45] For Dennett (1987/1998, p229), to look upon a system as an intentional system is to assume that it is rational. But the rationality assumed is the observer's, modeled on human thought. In a broader sense, "rationality" is simply the system following its own rules and goals rather than what human observers have defined.[46]

With regard to intentionality, one must distinguish artificial from natural systems. The intentionality and rationality of an artifact is specified by the designer, and hence may well correspond to that of the observer; but this cannot be presumed of natural systems. Intentional connections within an organism do not necessarily correspond to an observer's reason, logic, or expectation, any more than any other part of nature does. Intentional description should target the intentions and point of view of the organism—at least as human observers can understand them. We may describe intentional behavior in terms of logic circuits or flow charts, for instance, which seem to have an objective existence, apprehensible in third-person terms. Yet, these are but similes that refer implicitly to the observer's rationality, while ultimately it is the agent's own purposes and strategies we seek to understand.

One may be tempted to dismiss purposiveness altogether—as anthropomorphism, as merely reflecting what human beings experience and imagine (Ellis 2000, p41). Yet,

---

through assuming the system to operate on the basis of intentions imputed to it. The intentional stance involves assessing the goals of an agent in terms of one's own—or considering another entity as though it were an agent like oneself. This is crucial to predicting its behavior, but insufficient to understand in principle how a system can *be* an agent.

[45] When the intentional stance fails or becomes unwieldy, Dennett recommends falling back on the design stance. But this can only reliably be applied to *artifacts*, knowledge of whose design can be presupposed. Otherwise, we must fall back on the "intentions" of Mother Nature (natural selection) as designer—invoking the intentional stance again.

[46] Dennett (1987/1998, p234]) also claims that intentional theory "is vacuous as psychology because it presupposes and does not explain rationality or intelligence." But the circularity is Dennett's and does not inhere in the idea of intentional system.

taking the intentional stance in the case of many living creatures is simply common sense, provided we do not take such terms as purpose, belief, desire, and intention too literally in human terms. The intentional stance only appears as a special "stance" at all in contrast to the reduced kind of explanatory terms acceptable in physics.

5.4 Toward a theory of agency and internal communication

An agent does not have to be conscious, but one cannot understand consciousness without considering the internal agency of the organism, as well as its agency in the world. Both aspects must be understood in the context of the evolutionary and ongoing interactions through which an essentially syntactic system comes to make essentially semantic reference.

As presented here, the hard problem is not to reduce the mental to the physical, but to understand how an agent can have phenomenal experience. Such an explanation must not already presume the consciousness it tries to explain.[47] The strategy here is to take intentional connection and agency as fundamental concepts, define them apart from phenomenality or consciousness, then show how they can give rise to phenomenality and consciousness. I propose that the brain creates the "virtual reality" of conscious experience as a narrative to itself—as an efficient way for an executive function to monitor the world, the body, and their relationship, in the service of control. The question then becomes, why must this sort of monitoring be *conscious*? In other words, since many control functions are unconscious, why can't this be done unconsciously?

The hard problem rests on an intuition that matter and mind are two incongruous categories, if not different sorts of thing, so that our notion of physical matter cannot be made to account for our notion of consciousness. We do not readily see how a physical system can produce phenomenal experience. There is at our disposal, however, a contrary intuition: that a non-living artificial system (e.g., a computer or robot) *could* be conscious and enjoy phenomenal experience. These two fundamental intuitions are the bases of opposing positions concerning whether the hard problem can be solved at all and the kind of solution required. What is the basis of this second intuition? Why do we imagine such a possibility in the face of the other intuition? Surely, it is because we know that our own experience and behavior are intimately linked. If an artificial creature could *do* all that a human can do, then should we not conclude that it would also be conscious and have phenomenal experience? There are reasons to doubt that human behavior (or the structures responsible for it) can be exhaustively simulated or mapped. And there are those who doubt that even a functional duplicate of a human being would necessarily be conscious. But let us set aside these considerations for the moment. What informs the intuition is simply the notion that behavior and experience are deeply entangled. And that is the line of argument we shall follow.

If we can imagine that an artificial system could be conscious, it is because (rightly or wrongly) we believe it to be an agent in the sense that we ourselves are agents. Functionalists tell us that it is structure and connectivity that count, not the physical materials. We add to that principle the further idea that such a system must act on its own

---

[47] Nicholas Humphrey's theory, for example, has been criticized for containing key elements that already possess mind or consciousness (de Quincey, p79).

behalf, from its own point of view.[48] We understand how a machine can produce a virtual reality for a human user—because we have constructed machines to do this. The challenge is now to grasp how and why a machine could produce a virtual reality *for itself*. How does there come to be an actual view from its point of view in space and time, and "something it is like" from that viewpoint for the system itself? Ultimately, the task would be to write a virtual reality program for the brain.

The present theory is that, by its own internal acts and for its own purposes, an intentional system makes internal connections that it embraces as "true." This is how it creates a point of view of its own from which the world appears "real," independently of the point of view of human observers. In other words, an intentional system can communicate with itself by means of an internal language, and some of what it communicates *about* will be elements that appear in the observer's world. The meaning to itself of its internal communications is analogous to the meaning that emerges in the (human) act of reading or writing, of speaking or listening to a story or discourse. The human communicator's brain translates symbols (written or aural) into mental images, thoughts and feelings, or vice versa. The symbols bear no intrinsic meaning, but the language user *assigns* them meaning by making connections we have described as intentional. The brain is an intentional system whose *internal* language evokes its own phenomenality, in much the way that words evoke mental images for the human language user.

Just how does the brain do this? How do elements of an intentional system come to refer to, or represent, elements in the external world (Kuipers, 2007, p86)? My very short answer is: *by fiat*. An agent *makes* the reference, for its own use. The intentional connection has meaning to it because the object of the reference is significant to it. Reference (and therefore meaning) resides in such intentional connection, not in symbols themselves, which have no "pre-existing intentionality." There is no logically necessary connection between the sign and what it signifies. This internal "language" is not purely syntactic because the organism is motivated to establish reference outside itself, to the environment on which it depends. And that is the whole point: all meaning—including the meaning implicit in phenomenality—reflects the active relationship of the creature to its world. This creation of meaning is a process of internal communication about interaction with an environment full of consequence for the organism.

Just as squiggles on a page come alive as a story, and mathematical symbols may represent actual relations between things in the world, phenomenality is how the brain *represents to itself* its own intentional connections that bear significance for it, especially in terms of a putative external world. Phenomenality only occurs when there is need for explicit representation; otherwise, intentional connections are unconscious, though nonetheless meaningful to the organism.

The self-evident effulgence of *qualities* in sensation (the redness of red, the hurtfulness of pain) arises in much the way that the meanings of language do: as representations to the self. In natural language, sounds and symbols are *made* to carry meaning as words through a constructive process, just as algebraic symbols gain numerical significance by the mathematician's declarative fiat: '*let x stand for such and such…*' Phenomenal qualities that "emerge" in consciousness are comparable to

---

[48] An agent is not only physically realized but also *embodied*, which means it is an *autopoietic system* (effectively, an organism), or a function within one.

intelligible meanings that emerge through the babble of spoken syllables or the symbols on a written page. The relation between the phenomenal experience (such as auditory tone) and what it represents (such as longitudinal vibrations) may appear arbitrary and mysterious to an outside observer, just as the symbols of a foreign language appear mysterious to one who does not speak it (Grahek 2001, p150). We do not normally question the reasons for our own internal connections, however, any more than we question the use of familiar words. It is only from a third-person perspective that habitual connections appear arbitrary, because the observer is then not in the position of being the agent that makes them—even when that agent is one's own brain.

Yet, one may ask, to whom and by whom are such "meanings" conveyed? What intra-personal agents are involved in this internal communication? The question should be taken seriously, to open a legitimate avenue for further research. It should not be dismissed in the name of the Cartesian theater[49] or other projections by the human observer. While the organism as a whole is the agent of its molar behavior, neural processing cannot be understood without embracing (sub)agency within the organism as well. Otherwise, it can only be understood in terms of efficient cause—or following a rationality that belongs to the observer—which is to say, not understood in its own right at all. It might seem that ascribing agency to parts[50] within an intentional system does no more than regress the problem by assuming consciousness in the first place. My intention is rather to grab the bull by the horns and show how an agent or sub-agent can have a point of view and deal with meaning; and then to argue that under some circumstances representation *must* be conscious to serve the organism as a molar agent. In other words, intentional connection, representation, and meaning are necessary for both phenomenality and conscious action, though not sufficient. The task will then remain to show what makes the difference between the necessary and the sufficient conditions.

I hold that the creation of sensory experience is *like* the creation of mental imagery in response to language. Full-blown sensory images *resemble* their subtler cousins, mental images. Of course, simply pointing to a similarity does not explain what they do and do not have in common. Yet, their very differences may provide a clue to the sufficiency condition for phenomenal experience.

One difference stands out: mental images convey only the detail they already embody. Unlike a sensory image, a memory or visual imagining cannot be searched for more propositional information than it already presents, which is information stored internally from input up to that point. A retinal image, in contrast, is constantly updated in real time (or nearly), and is thus an ongoing source of new data. The visual field itself changes as the world changes, but is also continually refreshed through eye saccades. This constant renewal of an external source of sensory input (somehow) gives rise to a vivid experience of reality as distinguished from paler imagination and memory. (The question is: how?) But there is another factor than live input that is required for meaning, and therefore for consciousness; and that is the significance of the input to the organism.

---

[49] The idea that there must be a little person inside the brain to experience the "show" of phenomenality. Since the task is to explain "experience" in the first place, the Cartesian theater simply regresses the problem—a fact that Descartes himself realized.

[50] Cf. Sloman (2011, p28): "An information-user can have parts that are information users. This leads to complications such as that a part can have and use some information that the whole would not be said to have."

6.0 Embodiment, valuation, and affect

The organism perceives the world from the point of view of its needs and from its literal point of view in space. For many creatures, this includes perception of the body itself within that world, which serves as the reference to determine scale (Lehar, sec6.4). Perception of the world around the body would be useless if it did not relate to the body's needs and position in space.[51] The very idea of "reality" derives from effects upon and by the body in the physical world, and refers generally to the power of the environment over the vulnerable organism. Cognition is thus embodied by its very nature. And embodiment is more than just physical presence: it is 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 meaning to the organism of its internal communications refers ultimately to its evolutionary history.

What is the meaning of such "meaning"? The meaning of a proposition entails its significance to the agent that proposes it. Meaning is not necessarily something consciously communicated, but must refer to the potential consequences for the agent of events to which it is intentionally and causally connected. The physical reality of the world and the physical embodiment of the intentional system together underwrite the significance to it of external events. But significance also implies values according to which it is evaluated—particularly for implications to act.

To the fly, the descending swatter means "get away before you are squished (or eaten)". The fly does not need to have phenomenal experience, think, or make a conscious decision to perform this evasive action, since the reaction can be directly automatic (and indeed must be, for the sake of speed); yet the action expresses the intention to survive. The human with the fly swatter intends to kill the fly, but the fly does not need to know the human's intention in order to take the evasive action. (Presumably, it is all the same to the fly whether that is a fly swatter wielded with intent to kill or a hand merely chasing the fly away.) A social creature, on the other hand, often *does* need to read (imagine) the intentions of its conspecifics, who may even send deliberate warnings of their intentions as a substitute or prelude for more serious consequences. Either way, meaning refers to possible consequence and response, which presupposes a system of values (survival is good) and the possibility of effective action (quick take off).

Homeostasis is the basis of the valuation implicit in cognition. The creature must seek out external conditions that permit internal conditions to remain within tolerable

[51] Note that dynamics is kinematics plus force; and force is a concept that depends on interaction with the body, without which there could scarcely be a concept of nature at all (Jonas, p24). Notions of substance, force, mass, and inertia have an intuitive basis in body experience. These are precisely the notions taken up by Newtonian physics in distinction to the Cartesian physics of pure extension, and which have proven problematic to a science that excludes secondary qualities. Thus, definitions of force tend to involve circularity: mass is defined in terms of force and vice-versa. This is unavoidable when no common referent in body experience is allowed.

limits, and must reject external conditions that do not.[52] If able, it must act in such a way as to maintain favorable conditions. There is little point in something "having significance" if there is nothing you can do about it or if you don't care. On the other hand, "caring" (evaluating significance) doesn't have to be conscious. It is enough that those creatures survive that are programmed in such a way that they take appropriate action with regard to various stimuli. Valuation is required for true (autonomous) intelligence, and the *conscious* experience of that valuation is *emotion*. The fact that cognitive science has tended to ignore the affective states permeating all of human experience is a prejudice that derives from the exclusion of the subject, the body, and "secondary" qualities from physics.

6.1 The evolutionary source of meaning

The development of the organism's intentional connectivity is guided by natural selection, through which a physical system comes to be an intentional system. Such a system acquires a point of view of its own through a long evolutionary history as an intentional agent. It makes sense of its relations to the world through this history of interactions that matter to itself. Motivation is the source of all meaning, and motivation comes from embodied participation in an evolutionary contest: survivors are selected which are motivated to survive. The premises of the embodied mind are not arbitrary, programmed from outside or from on high, but inhere as values implied in genetic fitness. The origin of an organism's cognitive premises cannot be accounted for without an appeal to its evolutionary context, which provides the reasons for its reasons. It is not wetware that makes the brain different from a robot in its potential for consciousness, but its embodied motivations and evolutionary history, which give it "original intentionality."

The functionality of perception is a different issue than its correspondence to reality. Realism does not mean isomorphism of an internal model to external reality (resemblance, as with a painting), but conduct that permits survival in a context. The fact that our relationship to reality is thus indirect hardly implies either solipsism or idealism. It does imply that what is real for the organism is what can affect its well-being and which constitutes a potential reason for action (Beaton, 2009, sec 2.3.3). Truth and reality are idealized notions in the cognitive domain of the human observer, and not necessarily in the domain of the organism. They may not coincide with fitness, especially not when the cost of information is greater than its benefit.[53] If it seems an organism can survive despite an inadequate representation of reality, it is because fitness has little to do with the human ideal of truth.[54]

---

[52] The cell wall is the first boundary between internal and external, and the basis of the distinction between self and other—hence, of the distinction between what is good or bad for the organism (de Quincey, p72).

[53] Mark et al (2010, p504): "… truth can fare poorly [in computer evolutionary games] if information is not free; costs for time and energy required to gather information can impair the fitness of truth."

[54] Perhaps a modern analog of Descartes' faith that God is not a deceiver is the faith that Nature would not design us to have ideas contrary to our survival. After Darwin, the very notions of truth and error would be redefined in terms of adaptation and natural selection.

What is the relationship between evolutionary constraints and phenomenal experience? Only the behavioral concomitants of consciousness can enter into the mechanics of selection (Jonas, p127). It might seem, then, that nothing in the process of natural selection sheds light on the adaptiveness of consciousness or phenomenality itself (Horst, 1999, p44). However, natural selection can account for the organism's *valuations*, which are the basis in common of behavior *and* phenomenality. An organism's internal representations (where they exist) are not merely descriptive but are above all normative, serving to evaluate stimuli (Sebastien, 2016). They refer not only to something in the world but also to values and purposes of the organism. Intentionality serves control. It does not exist just to process "factual information" like a computer, without motivated implication for action (Sloman, 2011, p23), Without motivation, no mere collection of facts leads to intended action (Beaton 2009, sec5.2). Affect is the phenomenal experience of motivation.

Meanings in the lexicon of the organism's phenomenal language are established through natural selection. Clearly, pain *must* "hurt" if the creature is to survive; sugars *must* taste "good." Similarly, space *must* have a look of "depth" if visual appearance is to guide distance perception to accommodate movement through it. Above all, the world *must* have a solidly "real" look and feel to it, if phenomenality is to help negotiate its dangers and promises with due respect for consequences. Such appearances involve the organism's presentation to itself of behavioral concomitants implicit in its experience.

This is not to say that phenomenality necessarily or automatically implies behavior, or is required for it, but only to say that it is grounded in it. Much of phenomenal experience (for example, in the field of vision) is not directly related to survival or has any behavioral implication. And much of functional behavior is unconscious. Pain in mammals, however, seems to serve an alerting and motivating function beyond simple reflex, and favors protective behavior toward the injured part.[55]

Pain and pleasure are responses that obviously involve valuation. But what is the valuation involved in other qualia, such as color? Many species do not enjoy color perception, the functionality of which is particular to ecological niche. Humans inherited color perception from tree-dwelling primates who occupy a niche otherwise reserved to birds (who have keen and extended color discrimination). To compete with them, primates evolved color discrimination of a similar order (Humphrey, 1976). But discrimination is a form of behavior, which does not necessarily imply phenomenal experience. It remains to establish the nature and basis of the link between behavior and experience—and the specific role of consciousness.

6.2 Meaning and truth in perception

If the organism could do no more than receive physical energy, it would be logical to wonder how physical stimulation can be experienced as imbued with meaning (Dreyfus).

---

[55] In contrast, some insects—which lack specific nociceptors but may have pre-programmed avoidance and escape responses—probably do not experience pain or have any need to. In particular, they do not exhibit protective behavior toward damaged limbs or other parts, but stoically attempt to carry on as usual in the face of injury or missing parts. (Eiseman et al ,1984, p165). Nevertheless, the authors caution against jumping to conclusions.

The answer, of course, is that the organism does far more than passively receive energy from an environment. It interacts with an environment full of consequence for it and assigns significance and meaning according to its own values and criteria, which have been conditioned by natural selection. Its values reflect its evolutionary success, though not necessarily truth.[56] Its brain evolved to guide adaptive behavior in a specific niche and for specific purposes, rather than to produce veridical perceptions (Hoffman, 2011, abstract).[57] Color perception, for example, provides useful information at the human scale—such as which fruits are ripe. It is indirectly related to detailed microscopic properties of light (wavelength) and surfaces (reflectance). However, the question of whether or not the external world *has* color is misguided if not meaningless. It is a fact that the world has fruit and light and surfaces; but color involves the *interaction* of these with the perceiving subject. The same may be said of other properties—even so-called primary qualities such a shape, time, and "objectness."

Helmholtz originated the idea of the *perceptual hypothesis*—the progenitor of the modern idea of representation—and held that the senses provide only signs that must be interpreted, which constitute an inner language. What distinguishes this from the symbols of speech and writing is that their meaning has been established not by human convention but by nature (Helmholtz, 1867). According to Helmholtz, like scientific theories, our perceptual hypotheses are deemed true when they allow us to predict the results of our interactions with the world. What Darwin added to this picture is that—unlike scientific theorizing—the test of perceptual hypothesis is not experiment performed by conscious individuals but by natural selection. Perception is "veridical" to the extent it helps the organism to exist, or at least does not kill it. The same may be true of scientific theory. Of course, the scientific ideal of truth does not wait to see which ideas kill us, but seeks rather to predict the results of experiments and to underwrite technology. This departure from the evolutionary program distinguishes *truth* from fitness or (evolutionary) *adequacy*. It also contributes to the explanatory gap, since our ordinary perception does not correspond well with the picture of the world provided by science. There are conscious intentions behind the principles of science. While there are reasons for the specific characteristics of our phenomenality, they may not be *our* reasons but those of distant ancestors (Dennett, 1991, p381).

It would not be surprising if we share biases with other creatures with whom we also share most of our genetic makeup. Experiments with monkeys showing their responses to various colors shed light on some human preferences (Humphrey, 1976). They also hint at the evolutionary meanings of color qualia as "a species of affect-laden intentional activity" (Rovane, 2000, p89). Following Darwin, we should expect qualia to

---

[56] As one commentary wryly overstates it: "Perception is not about truth, it's about having kids" (Hoffman & Prakash, 2014, p3). But, if *all* cognition is merely adaptive behavior, is pursuit of scientific truth merely a strategy to have kids?

[57] Cf. Mark et al (2010, p513): "Fitness and access to truth are logically distinct properties. More truthful perceptions do not entail greater fitness… One key insight here is that perceptual information is not free. For every bit of information gleaned by perception there are typically costs in time and energy…A second key insight is that perceptual information is shaped by natural selection to reflect utility, not to depict reality. Utility depends idiosyncratically on the biological needs of each particular organism."

be functional for survival (Jackson, 1982). Far from being epiphenomenal, they should be intimately linked to appropriate behavior.

6.3 Phenomenality and behavior

The organism's responses are based on valuation, the phenomenal experience of which is *affect*.[58] While behaviors of aversion and attraction need not involve consciousness, consciousness necessarily is grounded in the values behind such behaviors. Feeling, with the judgments behind it, is therefore central to consciousness. If there is a primordial interaction of organism with environment, in which valuation enters as behaviors of aversion or attraction, then this will be the minimum situation that can invoke affect—and therefore consciousness. The situation with most immediate consequence is direct physical contact, and touch is the perception of it. Touch is the non-distance sense, in which information *about* the world is inseparable from appreciable causal effect *by* and *upon* the world.[59] At that primordial level, affect and its associated behavior form an integral whole—a notion with some history. Herbert Spencer (1890) had a basic intuition that the key to the nature of affect as a phenomenal experience lay in the behavioral response associated with it. He posited a primitive "shock" that is a kind of atom of feeling-response, from which the brain differentiates various sensations according to modality.[60]

The paradigm example is pain, which is at once feeling and response. Pain signifies and therefore *means* avoidance, or protection of an injured part. But the pain response has two phases, corresponding to two neural pathways (c-fibers and a-fibers).[61] One is a quick reflex reaction corresponding to the "shock"—removal of the hand in response to contact with a hot surface, for example. The slower response of lingering painful sensation reflects a persisting (internally generated) stimulus and acknowledgment of the tissue damage. Some version of it may persist throughout the process of healing, which takes place over time; the response associated with it is protective behavior. We might think of the ongoing nerve signal responsible for lingering pain as consisting of deliberate re-iterations of the first impulse—a reverberation or

---

[58] Cf. Dennett (1991, p176): "…the point of view of a conscious observer is… a sophisticated descendent of the primordial points of view of the first replicators who divided their worlds into good and bad." Valuation lies behind all intentional behavior, whether conscious or not, and so is necessary for consciousness though not sufficient

[59] Literally action and reaction. Cf. Jonas, p148: "… touch is the sense, and the only sense, in which the perception of quality is normally blended with the experience of force, which being reciprocal does not let the subject be passive; thus touch is the sense in which the original encounter with reality as reality takes place."

[60] Cf. Dennett (1991, p181): "This [evolutionary] history has left its traces, particularly on the emotional or affective overtones of consciousness, for even though higher creatures now become 'disinterested' gatherers of information… the innate links of informing states to withdrawing and engulfing, avoidance and reinforcement, were not broken, but only attenuated and re-directed."

[61] As Dennett points out, the physiology of pain is more complicated than this, involving separate channels through the "old brain" and the "new brain," and also the possibility of other pathways influencing the experience of pain. This should not affect the argument here, which concerns the grounding of the qualia of pain in the associated responses. See: Dennett (1978, p200-202).

reactivation loop persisting in time (Humphrey, 1992). The integration over time of these reiterations constitutes the *quality* taken to signify the persisting damage—namely, the hurtfulness of the pain. This conscious experience of pain carries several implications: first, that the initial reflex was not sufficient to avoid damage; second, that the injured part must be favored during healing; and third, the lesson to avoid such stimulus in future; fourth, that the experience is not a passive suffering but an active management of the situation.

On the basis of this kind of simplistic analysis, can we hope to understand, in a parallel way, qualities that do not seem to involve a reflex response or other associated behavior? Can we hope to understand, in terms of valuation and affect, such qualities as color,[62] auditory tone, smell, etc.? A reflex response to a proximal stimulus has immediate benefit for a creature. However, on the model of pain, it is not this immediate action but mediated *valuation* after the fact that is salient for qualia and upon which *subsequent* action is to be based. The re-action to a hot surface occurs before the onset of pain, which then involves an ongoing response to tissue damage. This evaluation response continues to carry with it the implied (but not actualized) reflex motor response.[63] The question is how to interpret, in terms of affect, qualia that do not seem to involve any action, let alone a reflex impulse.[64] It does not seem that there can be a comparable behavior associated with visible light of a given wavelength, for example, which is too weak a stimulus to directly affect the organism through simple contact (Jonas, p29). (For most wavelengths, there is little impact involved for the cell in an encounter with a photon.) Rather, the organism must *interpret* such a weak stimulus as having a significance requiring *considered* action—which is complex behavior, not simple response.

---

[62] Many organisms do not have color receptors at all, which in some creatures (such as birds, insects and primates) result from specific mutual adaptations with flowering plants. That is, color evolved *semantically*, first as a signal issued by certain plants and understood by certain animals; then as signals among animals, such as colored plumage; and now as artifactual signals used by human beings, as in lipstick and paint.

[63] There is neurological evidence that phenomenal experience involves efferent as well as afferent nerves. Cf. Ellis (2000, p44): "Consciousness occurs only with the efferent activation of the anterior cingulate, prefrontal and parietal areas, and this activation is not a direct causal result of primary or secondary projection area stimulation. This means that… consciousness does not result passively from causal stimulation… Consciousness always involves efferent activity, defined as neural activity generated by the organism itself, for purposes of its own survival and well-being, rather than from passive stimulation by incoming sensory signals." Cf. also (Humphrey, 2000b, p17-18): "In order to be able to represent 'what's happening to me', the animal must in fact continue to issue commands such as *would* produce an appropriate response at the right place on the body *if* they were to carry through into bodily behavior. But, given that the behavior is no longer wanted, it may be better if these commands remain virtual or… do not in fact have any real effects."

[64] Cf. Humphrey (2000a), who frames an account in the spirit of Spencer: "Now, it is true that, today, these sensory responses are largely internal, covert and private. But, or so at least I want to argue, it was not always so. Rather, these responses began their evolutionary life as full-fledged bodily behaviours that were unambiguously in the public domain — and, what is more, as behaviours with a real adaptive role…"

All nervous activity consists in the same sort of electrochemical signals. All extero-senses evolved from the cell membrane of the organism. The senses become physically differentiated, so that light, for example, is no longer (just) a proximal stimulus with an immediate import, but primarily conveys information from a distance. It no longer bears the same impetus of direct contact but engages a different level of response: monitoring and evaluating from afar. We have already noted that the distance senses are by definition freed from the need for immediate response. Visual and aural qualities are accordingly dissociated from such response. Yet, plainly, auditory tone results from repeated iterations of a wave front impinging on the eardrum. The single wave front (like the single photon) is a kind of shock, though involving minimal impact. The experience of tone emerges as the encounter with many wave fronts is synthesized into an experience representing an overall "texture."[65] Similarly, the visual world emerges from repeated saccadic "takes" on the retinal surface, without which blindness momentarily ensues (Solms, p183). And pain emerges from an ongoing reiterated stimulus that was initially but a shock causing a reflex response; without deliberate reiteration, the stimulus would diminish or disappear through adaptation. The brain integrates such micro-events into a gestalt by fiat. But to what end?

While unconscious mental processing occurs before the conscious experience it underlies, this does not render the phenomenal experience superfluous. Rather, the conscious experience indicates recognition, after the fact, by an executive function (the conscious self) of that particular unconscious processing. The conscious experience is a separate event from the reflex response, serving a different purpose with a different associated behavior. Consciousness is a form of monitoring of the outputs of various activities of the nervous system.[66] In the case we considered (of the hand withdrawing automatically from contact with a hot surface), the persisting experience of pain serves not to avoid the original stimulus, but to avoid further damage either by renewed contact or by lack of care for the damaged part. The pain as an ongoing experience marshals the conscious attention of the organism to protect the injured part and avoid further or future injury. It forces the organism to favor that part in order to facilitate a healing process that takes time.

A similar divide between quick and slow pathways seems to exist in the visual system too, where an initial fast wave of visual processing happens outside consciousness, but is made available to subsystems for immediate reflex responses. This is followed by a slower phase of "recurrent processing" that involves integration of various brain areas and may lead to conscious experience (Revonsuo, p215, citing Lamme). Here too qualia emerge from reiterated signals and serve a different purpose than the initial signal.

Just as there may be ongoing pain, the fact that there are ongoing visual and other qualia reflects the fact that the world bears significance for the organism on an ongoing

---

[65] Jonas, p29: "The smallness (in dimension, time rate, and energy) of the unit-actions and reactions involved in affection of the senses… permits their mass-integration into one continuous and homogenous effect (impression)… Where qualities are perceived, the raw material is action: impacts, hustlings, clashes on a molecular scale. Organisms not far exceeding that scale can therefore have no perceptions, but the collision experience [sic] only."

[66] Cf. Humphrey (2000b, p13): "So that what I actually experience as the feeling—the sensation of what is happening to me—is my reading of my own response to it."

basis. While many responses can be handled unconsciously, it serves some organisms to monitor their sensory input for events that potentially cannot be handled by automatic behavior. Conscious experience is the first-person version of that monitoring, by a molar agent with executive responsibilities, which may include "broadcasting" pre-conscious information to various sub-agencies (Kandel, 2012, p464-5).

6.4 The language of the senses

What the color red signifies to the organism may not be the same as what it did a million years ago, just as archaic English words do not have the same connotations to a modern ear as they did to the Elizabethan court. Shakespearian scholars attempt to make up this deficit, and evolutionary scholars attempt to unravel the changing meanings of the language of the senses. The intentionality involved in color perception is deeply buried in the evolution of the primate visual sense. We have little current need to be engaged by past behavioral correlates of color qualia, first because those associations have changed over time; and second, because the nature of the visual system is to divorce itself from immediate behavioral implications altogether. We gain objectivity at the price of access to the underlying subjectivity. The problem presented by the seeming objectivity of the visual sense (and perhaps that of science) is to understand how it derives so indirectly from affect.

Can one then speculate on the "meanings" of colors? Evolutionary advantage readily explains color *discrimination*—why things appear to be colored differently—but not why a particular wavelength of light is experienced just so and not otherwise, nor why colors seem to be grouped in categories. The very existence of color categories (hue) reflects needs of the organism more than just properties of light or reflective surfaces.

While the "primary" colors are idealized categories, they do correspond with particular sensitivities of receptors. Trichromancy enabled primates and their human descendants to distinguish effectively in a range of wavelengths corresponding to phenomenal red through phenomenal blue. As Newton had noted, red and blue are subjectively connected at the farthest ends of the spectrum, so that violet seems to mix blue with red. This forms a cycle (hence the color wheel) and falls just short of covering an "octave" of wavelengths.[67]

Color vision occurs in some fishes, reptiles, insects and birds (Humphrey, 1976). Human beings have color vision because they are primates, which evolved it to occupy a common niche with birds. The diets of Old-World primates consist significantly of fruits that are yellow, orange or red (Tsou, 2013). It would make sense for these food items to stand out from a background of foliage, and subjectively we think of red as contrasting maximally with green. (Yellow is closer to green in wavelength, and also may indicate a

---

[67] That is, the red end of the visible spectrum is just less than twice the wavelength of the blue. One could speculate that, if human vision extended over more than one such octave (i.e., into infrared and ultraviolet and beyond), the subjective experience would resemble that of sound octaves (with pitches at multiples of each other appearing qualitatively similar though higher or lower). This raises the question, however, of how the hue in one octave would resemble or differ from the "same" hue in another. On the other hand, some tone-deaf people cannot tell which of two auditory tones is "higher" in pitch.

less ripe fruit.) In the forest context, at least, the color red serves to alert the creature to something singular—whether a ripe fruit or a poisonous creature that has adopted the color code (Dennett, 1991, p385). But there would be other associations possible, such as blood, or colorings related to sexuality, or the red of dusk when some predators hunt. No doubt all qualia involve a network of such associations, for which the quality itself stands in summation (Loorits, citing Crick & Koch). [68]

This still does not tell us why the chlorophyll of the forest does not appear red, and the ripe fruit green, for example, which would maintain *the same contrast* for discrimination. What is it about the qualitative feel of greenness that commends it to represent foliage in the vocabulary of the senses, and what about redness commends it to represent things that must stand out in the forest?

This question is like asking why a particular meaning is denoted in the English language by a particular word,[69] written and pronounced its given way, rather than by some other symbol or sound. For the native language user, the association seems natural and unquestionable, though of course it is logically arbitrary and a product of historical accident. The subjective experience of qualities—in this case color—arises from sensory input in a way analogous to how meaning arises from the sounds or characters of language (Bruiger, 2024). The internal language of the organism may be no less arbitrary and accidental in its choice of symbols than English. *Some* symbol must be chosen, which will inevitably come by convention to seem imbued with the meaning it has been made to convey by virtue of its connection with the real world. So, it is backwards to ask why grass "feels" green; rather greenness is imbued with the association of grass. Greenness is the way we visually experience the totality of associations related primarily to chlorophyll.

But language is also *motivated*. Originally it may have served to alert or warn. Today one may discourse dispassionately about many topics, but the first human vocal expressions were probably not so different from the excited alarm calls of primates and other animals. What makes fully grammatical language such an invaluable tool is precisely that it may be used in a detached and flexible way, combining words and ideas in inventive new permutations, removed from their original, or any specific, context or

---

[68] Cf. Dennett (1991, p378): "Once there were creatures who could distinguish red from green berries, they could also distinguish red rubies from green emeralds… Why is the sky blue? Because apples are red and grapes are purple...." But, why would the distinction between red and green imply other color representations? Dretske (1994) has an answer: Once a scaling structure has been established, there is no need for each representational state to be separately evolved. Once "12" is taken for midnight or noon, the other positions on the clock face are implied. That makes sense for numbers, whose relationships are formally defined. But what is the formal relationship between colors? Once "red" and "green" are associated with certain natural features, how does "blue" *follow*? Cf. also Humphrey (1976): "We may presume that colour vision has not evolved to… see simply the greenness of grass or the redness of raw flesh, since those animals which feed chiefly on grass or on flesh are colour-blind… But the most striking colours of nature, those of flowers and fruits, the plumage of birds, the gaudy fishes of a coral reef, are all 'deliberate' evolutionary creations which have been selected to act as visual signals carrying messages to those who have the eyes to see them."

[69] The simile is not new. "For Cudworth, like for Descartes, sense can be compared to speech by which 'Nature as it were [talks] to us in the sensible objects without, by certain motions as signs from thence communicated to the brain'… (Mausfeld, 2002, appendix)."

need. Vision, especially, is detached in a similar way. The informing significance of most qualia is no longer compelled by their original urgent associations, from which they may have emerged in the way that grammatical language emerged from animal calls.

Certainly, some colors have affective values for human beings and for their primate relatives. They can serve to capture attention, to convey information, and to bear an emotional charge. Monkeys tested for color preferences showed a preference for blue and green and an aversion to red, which they found upsetting (Humphrey, 1976).[70]

In any case, the *sensation* of redness (unlike the word) is not merely a linguistic convention subject to social change, but a convention of neuro-logical organization, with the enduring force of long genetic precedent. Indeed, the human cognitive system gradually adapts to distorting colored lenses or filters in such a way that experience of verdant foliage, for example, is restored to normal greenness (Neitz et al, 2002). The words of a natural language have relatively transient reference, on the cultural time scale. The meanings of qualia are backed rather by evolutionary history. The sensation of redness is inherently different than the sensation of greenness precisely because of the real-world things it refers to in our evolutionary history, from which it cannot be arbitrarily dissociated. There is no "inverted spectrum" because human beings share a common physiology and "language of the senses."[71]

Differences among qualia in a given sense modality might be described abstractly by finding a ground for them, perhaps as information. Similarly, for differences between modalities. After all, whatever their source or destination, nerve impulses consist of electrochemical signals that the brain is able somehow to differentiate. Qualia could then be reduced to differences (quantities) among that common coin of information and among the structures that receive/emit it. But this sort of reduction is from the point of view of the observer. From the brain's point of view, qualia are precisely *how* it differentiates—and represents to itself—the signals that appear (to the observer) as nothing but different patterns of energy.

7.0 Having a point of view

The potential vista from the point in space and time occupied by an observer is different from what can be seen from any other vantage point. But what is actually experienced at

---

[70] However, as Humphrey explains, the laboratory context must be taken into account, where the natural implication for response cannot be established. It may be precisely the ambiguity of red as a signal that gives it power, and creates anxiety in a situation without a natural interpretation: "Red toadstools, red ladybirds, red poppies are dangerous to eat, but red tomatoes, red strawberries, red apples are good. The open red mouth of an aggressive monkey is threatening, but the red bottom of a sexually receptive female is appealing. The flushed cheeks of a man or woman may indicate anger, but they may equally indicate pleasure. Thus the colour red, of itself, can do no more than alert the viewer, preparing him to receive a potentially important message; the content of the message can be interpreted only when the context of the redness is defined. When red occurs in an unfamiliar context it becomes therefore a highly risky colour."

[71] The 'inverted spectrum' is the "apparent possibility of two people sharing their colour vocabulary and discriminations, although the colours one sees… are systematically different from the colours the other person sees." (Wikpedia: inverted spectrum).

that point will be different as well for any other observer that could occupy it, to the extent that cognition differs from individual to individual. "Point of view" therefore includes what the subject brings to the table, even while it still alludes to vision as the most "objective" of the senses.

Point of view is also a linguistic concept, expressed in the "person" in which statements are cast. This refers to who is speaking, who is spoken to, and what is spoken of. Typically, personal statements use the first-person pronoun to identify the speaker, thereby (at least inadvertently) claiming responsibility for what is said. Second-person statements can take the form of commands or requests. Scientific discourse is typically cast in third-person statements; the speaker and audience are taken for granted, to focus on the topic of discussion. This practice is justified by the presumed training and expertise of speaker and listener, their common terminology, and the protocols that are supposed to eliminate subjective factors from scientific method and discourse. The observer is standardized, her effects on observation minimized or neglected.

7.1 First and third person

The principal difference between 1st and 3rd person perspectives is that the former consists of phenomenal experience and the latter consists of facts derived from it or statement about it. Concepts belong to the 3rd-person realm, qualia to the 1st-person realm. On the other hand, phenomenal experiences (qualia) *implicitly* present information that might be expressed 3rd-personally (e.g., through scientific instrumentation). An organism's information space is partitioned between environment-related and system-related processing (Metzinger, 2005, p26). Properties of external things are at least indirectly referenced in phenomenal experience. However, qualia represent interactions of organism with world, not objective properties of the world independent of the organism. That is, they also reference properties of the organism or its representational system.

The physicalist program is to explain consciousness as a function of the brain, from whose point of view the world is external to the skull. From the point of view of the organism as a whole, the world is external to the skin. But from the epistemic subject's point of view, brain, body and skin are alike external to the mind's eye—as are *all* objects of perception, thought and imagination. The source of the subject-object relationship is this circumstance: that the whole realm of experience is seemingly external to the subject. In any case, point of view is more than a function of language or cultural assumption. It reflects the direction of attention. From a traditional scientific standpoint—with its realist direction of attention "out there" *at* the object of perception—the hard problem is the very existence of "in here": the seat *from which* perception takes place. In other words: the hard problem asks why and how there is any such thing as the first person (MacLennan; Frith & Metzinger). In that sense, the explanatory gap is between first- and third-person viewpoints. The problem is not that the gap cannot be bridged, for the mind/brain normally *does* bridge it in daily experience. The problem is to understand how it does this.

One version of the challenge is to show how 1st-person experience results from processes that are typically framed in the 3rd person. However, circularly, everything that is framed in 3rd-person terms results ultimately from 1st-person experience! This is no less so in the domain of scientific description, which implicitly considers its objects from the shared perspective of scientists, without regard for a point of view of the objects themselves. Historically, the western tradition has considered even living matter to be inherently inert and passive, in contrast to the active agency of mind. Because of this dualism, if for no other reason, there can be no adequate theory of mind until the viewpoint of the cognitive system itself is considered.

Scientific description is ultimately penned from the 1st-person point of view of the scientist, in 3rd-person language, communicated to other scientists. For the reflexive language user, three domains of description are thus involved: that concerning one's *experience* as a subject, that concerning the *objects* of one's experience, and that concerning the *descriptions* themselves as communications to others. "Others" can include oneself as recipient; and "objects" can include other persons and creatures presumed to be subjects as well as objects.

Scientific description is *communication*—in the *second* person—intended primarily for other scientists. It is not a snapshot of objective reality, taken from nowhere in particular. First-person experience is necessarily involved in garnering sharable facts about the world—through observation, measurement, and experiment.[72] There are protocols to govern observation so as to make observers interchangeable in principle; and there are protocols to standardize experimental procedure and communication. By definition, however, the observer's experience is not literally impersonal.[73]

One gives a first-person account when describing one's own experience as such, in contrast to describing the world. Using the first-person pronoun claims responsibility for the statements of the description. Even when responsibility is not explicitly claimed, such statements are acts of communication made *to* a recipient—even when only talking to oneself *about* one's own experience.[74] Communication is by definition interpersonal, and third-person accounts describe the world as others (presumably) can access it. This applies even to accounts that present to others facts *about* one's experience, since facts (descriptions) can be shared even if the experience itself is unique to the subject. The *fact* that something is red, for instance, is propositional knowledge that can be shared, even though it is not the redness itself. They are different domains of knowledge.[75]

Thanks to the computational metaphor, it is easier to identify neurological processes with rational thought processes than with sensations (they are both "processes"). Thought has a communicable propositional content, which phenomenality

---

[72] Even if the measurement reading is automated, at some point human (or other) consciousness must be involved in giving it meaning.

[73] Despite the "block universe" of physics, there is no gods-eye view even for the scientist. By making the first and second person aspects of scientific communication tacit, however, can give the misleading impression that the third-person stance is not a human communication but a transparent window on the world.

[74] Facts that one may introspectively note about one's phenomenal field, for example, are propositions—potentially sentences in language—if only communicated to oneself.

[75] One can know "everything there is to know" about Paris (through pictures, history, eyewitness reports, etc.) but it is not the same as being there. Also, one can know everything that another person can convey about their experience, but this is not the same as being that person or having their experience.

appears to lack. Propositions can be consciously derived *from* phenomenal experience (as in description), but any propositional reasoning *leading to* phenomenal experience is unconscious and we have no introspective access to it. This disconnect gives the impression of a qualitative "remainder" that resists explanation—the "something it is like" for the conscious self to be having a phenomenal experience. (Needless to say, there is also something it is like to be having a thought, only this is not considered germane to its propositional content—the part intended for communication.) The conscious self's lack of access to those brain processes that lead to its own experience is the counterpart of our lack of access to the phenomenal experience of others: in both cases it is a matter of physiology.[76] And in both cases it is language that has compensated for this lack. By means of language, we can speculate and communicate about other people's experiences, and perhaps feel empathy for them. We can also speculate and communicate about the "neural correlates of consciousness"[77] and how brain processes might produce phenomenality.

7.2 Conscious versus unconscious[78]

A great deal of human behavior is performed without conscious attention. We experience *results* of "neural processing" but not the processing itself. Some of this processing leads to conscious experience and some not. What makes the difference?

Conscious attention seems to be required in novel, complex, or otherwise mobilizing situations. If the action cannot be done by rote, if it confronts a novel or demanding situation, or if it requires planning and forethought, then conscious attention comes into play. This suggests that phenomenal experience makes real-time sensory input available to higher centers for planning or dealing with novelty or emergency. One role of consciousness, as a state of the system, is to muster additional resources to deal with situations that are not already handled by existing automatisms. The conscious state is called upon to gain new mastery.[79] Formal processes of reasoning can be precise, but are serial and slow compared to informal estimation or guessing, which is in parallel and much faster but probabilistic (Singer, sec 4.2).[80] Complex problems involving many

---

[76] One can experimentally have 3rd-person access to the neural processes leading to another person's experience, but not to the experience itself. In the unusual case where the subject and the observer are the same person (say you are examining a real-time MRI of your own brain), you would observe the MRI, which would be a read-out of your experience of observing the MRI…and so on, recursively.

[77] A neural correlate of consciousness (NCC) is defined as "the smallest set of brain mechanisms and events sufficient for some specific phenomenal state" (Revonsuo, p209, citing Crick & Koch,1990). This definition assumes some degree of localism, perhaps inspired by the computer metaphor where information has a specific address. If the "smallest set" includes the whole brain, NCC would be a meaningless idea.

[78] Properly speaking, 'unconscious' means a temporary condition, whereas '*non-conscious'* should be used to mean a permanent condition (Revonsuo, p89-90). However, I will stick with the popular usage.

[79] Effectively, to restore a state no longer requiring consciousness! (Solms, p179; Lamme, sec11).

[80] Cf. non-conscious visual perception (blindsight), which shows experimental results that are better than random but far less than certain.

variables, or requiring a quick response, are better handled by unconscious processes, which operate massively but give only rough answers (Kandel, 2012, p468). Conscious problem solving is slow because it involves many steps, with limited working memory, and coordinates more areas of the brain that are farther apart. Conscious visual processes, for example, seem to involve information shared over a key distance of 10 cm in the human brain (Lamme, sec9).

Consciousness plays a different functional role than behavioral responses that can occur without it, and presumably involves different neural processes. Many motor tasks are executed "automatically;" in some situations one is conscious *of* the action, which nevertheless seems to "do itself." Awareness of initiating voluntary activity comes slightly after the neural processes that have physically caused it, yet this awareness serves as the basis for choosing *future* action, or action in a larger context (Frith & Metzinger). Thus, consciousness is not epiphenomenal, but plays a specific causal role.

The conscious experience of estimating or guessing summons a consciously accessible answer from unconscious processing. Walking or driving a car does not require conscious guidance if the route is well known—as somnambulance dramatically demonstrates.[81] On the other hand—though they may demand no immediate action—planning, forethought, and reasoning obviously involve conscious attention, as do alarums and emergencies that mobilize the body and do require immediate response. Such facts suggest that consciousness serves an executive function and monitoring role, with much behavior or processing performed unconsciously by sub-systems that must be coordinated. Consciousness seems to involve a synchronous broadcasting of information globally, especially to parts of the parietal and prefrontal cortices (Kandel, p464-5).

The agent in this role is known to itself and others as the "self" of that individual, whose job entails a first-person perspective (Metzinger, 2010, p29). Though it can scarcely take full responsibility for the actions of the organism as a whole, and is certainly not identical with it, it seems to appropriate that identity to itself and is often held accountable by others. It is more than a figurehead but less than in full control. It is a sort of virtual representative of the organism tasked with specific responsibilities, which include monitoring a real-time virtual-reality version of the external world, which includes the organism's relationship to the world. This implies an interface, at which conscious experience takes place, between this agent and other (sub)agencies within the organism. Consciousness is also closely bound up with memory (Lockwood, p84). The monitoring function serves to register information coming into the system, tagging it for future retrieval.

Yet, one could still ask: Why could this not all go on unconsciously? Why could there not be an *unconscious* executive function, with *unconscious* monitoring, just as there can be unconscious representation or perception?[82] I believe the answer resides in

[81] Consider the case of a sleepwalker who (asleep) drove to a familiar house and murdered one of its occupants, with no apparent motive. He was subsequently acquitted on the ground of not being consciously responsible for the act [PBS documentary, "The Brain", by David Eagelman].

[82] Indeed, Lamme (sec2) raises the specter of unconscious *sensations*: "There are situations where it is in fact unclear whether we should talk about a conscious sensation or not." He gives the example of the split-brain patient, who may *tell* you that the stimulus was not perceived but is nonetheless able to *draw* it. This has sparked debate about whether consciousness can be

the nature of the executive function as hierarchical decision maker. Incoming sensory information is essentially probabilistic and ambiguous; only if deemed noteworthy is it registered as "received." As in a corporation or government, sub-agencies passing such information along to the executive agent must package it appropriately for decisive action. This means that what the conscious self (the executive) perceives must be definite and *un*ambiguous, even at the cost of being wrong.[83] One could say that consciousness reflects the definiteness of the digital, whereas the unconscious is probabilistic, vague, and analog. This by itself does not yet explain consciousness, if we are looking only from a third-person point of view. Ultimately, we must take the position of the agent involved, who *asserts* definiteness through decisive acts of fiat, which may send a kind of emotional alarm signal that upgrades the system from impassive monitoring to active attention.[84]

7.3 User interface

Quite apart from the question of consciousness, the interrelationship among various stages of cognitive processing may usefully be conceived to involve interfaces between them, with relative digital or analog aspects. The question is, then, what is unique about the interface we call conscious experience and about the agent who makes use of it?

A useful (though potentially misleading) feature of the computational metaphor is the notion of the user interface. On that analogy, conscious perception is a user interface between a particular agent (the self) and pre-conscious levels of processing. The events the conscious self registers have already been caused in the nervous system. But this does not negate an active role for it to play, any more than the hidden operations in a computer negate a role for the human user who responds to them. Quite the contrary, phenomenality is a display for informing high-level decisions that may then have a downward causal effect on the operation of the system. It is a phase of action, among others that remain inaccessible to consciousness.

A computer-generated image displayed on a computer monitor is useful to the human user as a read-out of processing within the computer, to summarize operations already performed by the computer in order to enable the user to direct it to perform further operations. From the point of view of the *computer*, of course, none of its operations require literally searching a graphic display, which exists solely for the benefit

separated from verbal reports—about the difference between seeing and knowing, or between phenomenality and access consciousness.

[83] Cf. Frith & Metzinger: "But only misrepresenting the probability of a hypothesis as 1.0 and simultaneously flagging it as a fact holding *now* via a window of presence turns a possibility (or a likelihood) into a reality. This is… the functional mechanism that 'glosses over' subpersonal Bayesian processes by assigning 'realness' to them—that is, by misrepresenting them as exemplifying *absolutely* maximal likelihood… Cf. also (Kouider et al, p303): "[With a degraded stimulus]…observers cannot report its total content but nonetheless experience a rich visual experience. This has been described as a form of cognitive illusion, whereby subjects have a feeling of seeing that does not correspond to what they can access." At the other end of the scale, blindsight is a situation where information can be extracted without (reports of) awareness.

of the human user.[85] By analogy, the relationship between the conscious self and unconscious processing is like that between human user and computer. While that's a metaphor, in crude terms it seems that the front half of the brain *does* serve as "homunculus" and executive agent to the back half (Hirstein, p153). Furthermore, no Cartesian theater or regression of witnesses is implied.

Nevertheless, the interface metaphor can be misleading in other ways. An *icon* on the desktop of a computer monitor, for example, represents a file or program in the computer, which can be activated to result in the sort of monitor display we have been considering (Hoffman 2008). Its appearance on the desktop is a higher-level (or meta) display, which represents a *potential* user interaction. It is not a read-out of the interaction itself (as when the program or file is opened); much less does it represent some aspect of the real external world. It is more like an abstract thought than a perception.

Another analogy likens the situation of the brain isolated inside the skull to a flight simulator, which nevertheless represents and controls a real plane in real time in real flight. Like the chamber of the skull, the simulator chamber is a windowless cockpit, with electrical input from various sensors (other than direct video), so that the pilot is flying entirely by instrument. The "picture" that emerges in the mind of the pilot is a matter of inference from the readings. This "pilot" has never been outside the cockpit to experience the external world directly. The brain is literally flying blind, and the experience might be compared to that of congenitally blind people who claim to "see" what is before them. The pilot must accept as a "true" representation whatever picture the simulator presents, so long as it doesn't result in crashing.

Though there is a witness, there is no regression. For, the sub-agents supplying information to the executive function do not recreate a version for their own use, in miniature, of the display intended for the executive. They have no need to, for they are not tasked with the specific executive responsibilities of the conscious self. Though there is a display, it does not take place *in* a theater. Rather, it *is* the theater: a virtual representation of the world, with a virtual representation of the self within it (Metzinger 2010, p34-35).[86] The experience of selfhood and will are the human first-person versions of a control function to guide the molar action of the organism. The self's powers as "captain of the ship" may be limited, but it is no mere stowaway.[87] Being captain does not mean micromanaging every operation involved in running the ship, only holding a certain executive authority.

8.0 Fiat: toward a theory of qualia

A solution to the hard problem must include a theory of qualia, which cannot be dismissed as "epiphenomenal" with regard to associated cognitive behavior or neural processing. In looking at a source of green light, for instance, one should not imagine that

---

[85] The information responsible for the pixels of a video image is stored digitally in memory. The computer can access and use that information without consulting the video display per se.

[86] What makes this virtual production more than a work of solipsistic fiction is the fact that sensory information is continually updated in real-time interaction with the world, as in "predictive processing" (Hohwy, 2013).

[87] PBS documentary, "The Brain", by David Eagelman

the brain assesses a frequency and then pointlessly "colors in" that information with the superfluous quality of greenness. Rather, the experience of greenness is how the executive function represents that frequency (and perhaps other information) to itself. The quality comes with the information built in. Similarly, the perceived quality of a particular musical tone is itself an estimate of sound frequency (perhaps among other things). Sensory qualities are thus not something above and beyond the information they represent, nor are they caused by the processes transmitting it. Rather, they are a *version* of that information, a synoptic way of presenting it to the executive function. Qualia are how the embodied subject first-personally detects properties that might otherwise be detected with laboratory equipment, and which an observer would describe in third-person physical terms. If the specific quality of greenness seems to convey privileged information beyond that involved in the public analysis of light, that is because (following the principle of co-determination) it also bears information about the organism itself and its relationship to the world. This includes information about its internal communication that mediates that relationship, its priorities and evolutionary history, and the sensory modality. In other words, the quality is both message and medium.

Propositional knowledge concerns differences discerned among qualia, construed as differences in the world. Such knowledge is *based* on phenomenal experience, but is not the experience itself. However, qualia themselves *do* contain implicit information about differences in the world, which propositional knowledge renders explicitly communicable. Qualia seem ineffable because they already and pre-verbally recognize what can be otherwise expressed in language. Just as propositional knowledge arises by conscious acts of assertion, so the qualia of first-person experience arise by unconscious acts of assertion (fiat).

Phenomenal experience does not seem to happen inside the skull, because what it refers to is not inside the skull. The whole point of perception is to navigate an external world, not to navigate the brain.[88] This projective capacity can exist because the relation between intentional connections and neural connections is *like* the relation between the meaning of what you are now reading and the paper and ink (or electronic pixels) that physically convey it. The ink is on the pages of the book, the digits are in the computer; but the story unfolds somewhere else.

8.1 Painting by numbers (*pace* Dennett)

Vision is the sense that human beings most rely upon to reveal objective properties of the world. Properties such as taste, odor, and even color, however, clearly depend as much upon the body of the perceiver as upon things in the external world. Through the visual sense, as through hearing, we are able to monitor the external world continuously and with the relative detachment that comes with the security of distance. The sense receptors of these modalities are modifications of the skin, and hence of the direct contact of touch. But whatever alarm is involved in direct contact is correspondingly mollified owing to distance and the relatively tiny impact of light and sound waves. Hence, visual and auditory sensations are not normally experienced as bodily sensations at all. For these

---

[88] The latter would be an interesting second-order natural ability, for which evolution has not equipped us, perhaps for reasons of economy. An artificial mind might incorporate this ability.

modalities, contact with the sensory surface is not experienced *as such*, but projected as an event in external space.

If there is a primordial response of the organism to sheer contact, the greatly elaborated responses of the distance senses will be adaptations of that to provide for continuous monitoring. Consider the quality of a musical tone. To be perceived *as* a tone, there must be a regular repetition of wave fronts hitting the eardrum. A single wave front impinging on it will not be experienced as a *tone*, if it registers as an experience at all. It will be an instantaneous disturbance of some sort, but not a quality. Similarly, the eye requires a certain number of photons to register anything at all, and more to register color. Paradoxically, it must constantly move (with saccades) in order to register a stable ongoing visual field. Without this continual shifting of gaze, so that information is constantly updated on the retina, the visual field dissolves. The succession of saccades provides an effect like the succession of wave fronts on the ear. The eye/brain smooths over the discontinuities of the saccades like it does the separate frames of a motion picture. Otherwise put, it "fills in" the continuity of the visual field like it fills in the blind spot and fills in the quality of a tone, integrated between successive wave fronts. This "filling in" reflects the intentional system's *assertion* of continuity, by disregarding an actual discontinuity. If the brain can create such an appearance by simply asserting that it is so, then why not all appearances? If it can project events on the retina or eardrum as taking place in real external space, then it can create the appearance of a real external world by fiat.

Qualities emerge in first-person experience through a process that seems to be a matter of "filling in" or "completion" of what would appear to a third-person observer as mere structure, fact, or data. (Let us recall, however, that such data, facts, or structure were abstracted or extracted from phenomenal experience in the first place.) I use the term *filling-in* advisedly, in view of Daniel Dennett's well-known critique of the notion.[89] His valid point is to distinguish between representing a continuous phenomenal field and representing *that* it is continuous. It is precisely this operation, of representing *that*, which forms the normal basis of qualia, though only in anomalous circumstances do we even notice it. These circumstances include laboratory studies of perceptual completion effects, habituation, perceptual adaptations of various sorts, and phenomena of spatial projection—all of which are examples involving acts of fiat. In such processes, features of the subject's phenomenal experience typically seem to go beyond the facts or events noted by the observer. That is, the subject's brain "fills in" what is "not really there" for someone in the role of third-person observer. Tones, colors, smells and qualities in general may be understood as such completion or projection effects, with the caveat that what is deemed "really there" involves a fundamental bias deriving from the scientific use of the third-person point of view.

The terms 'filling-in' and 'completion effect' are somewhat ambiguous, since they sometimes refer to the effect *experienced* and sometimes to what the brain supposedly *does*. There is evidence of a neurological basis for filling-in in the latter sense, compatible with Dennett's contention that the brain *ignores* an absence (Pessoa et al.). Yet, there is certainly something it is visually like to be doing this ignoring. The experience of a continuous visual field, in spite of the blind spot, is an experience of

[89] Dennett (1991, p356): "The fundamental flaw in the idea of 'filling in' is that it suggests that the brain is providing something when in fact the brain is ignoring something."

continuity, not an absence of experience.[90] I hold that qualia *in general* are a result of the same process as involved in the blind spot and other completion effects, which is precisely a matter of jumping to conclusions through positive acts of assertion. One might then ask, in the general case, what does the brain fill in *between*—as when, in the case of the blind spot, it fills in between the enervated retinal areas on either side of the un-enervated area? The answer must be that these areas too are filled in, but on a finer scale (between receptors, for example) and temporally as well as spatially, in the way that the brain fills in continuity between frames of a motion picture.

The overall conclusion is that the brain treats the phenomenal field digitally, so to speak, in the sense that discrete spatial or temporal areas are *defined* to exist, with the area lying within the defined boundary labeled as a quality—much as in "painting by numbers." If there is further differentiation within an area, it reflects the same arrangement on a finer scale. What is experienced in all cases is what the brain *decrees* to be there.

According to Dennett there is no "mental paint" to fill in the bare bones of propositional information in the nervous system, so as to wash it with "qualities." Dennett's very point is that the filling-in experience (as in the blind spot) is the brain's judgment that the missing information is *not important*, since it does not represent a hole in the *world*; the brain translates that as the experience that there is no hole in the *visual field*. What to call this act of interpretation 'mental paint' or 'no mental paint' seems no more than quibbling. In the case of painting by numbers, there is no finer detailing within a discrete area (no further information), which is what makes it "digital" rather than "analog." The coded number and the corresponding color are simply two versions of the information pertaining to that cell.[91]

8.2 Filling in and projecting out: phenomenality as completion effect

Regardless of how complex the neural processes giving rise to them, qualia are essentially simple integral gestalts (Crane, p188). This is so first of all because there is no conscious access to the differentiated structure of their underlying processes (Loorits); but equally because qualia are by design interpolations or syntheses that function to summate information.[92] Insofar as the brain as a virtual machine can only deal with

---

[90] "Experimental investigations of the blind spot, stabilized images, illusory figures, and chromatic diffusion and the gap effect provide suggestive evidence about certain consequences of perceptual completion. In particular, they point to measurable effects that seem to depend on representing a presence rather than ignoring an absence…" [Pessoa et al, p741]

[91] In the case of the blind spot, there are no sub-areas, only a uniform absence of input. The brain simply jumps to the reasonable conclusion that similar information is received within the blind spot, by interpolating with the surrounding areas. This is reasonable because the brain "knows" that the blind spot is a hole in the visual field and (probably) not a hole in the world. How we experience the result of that assumption is that the blind spot is "filled in" to be seamlessly continuous with the rest of the visual field.

[92] Jonas, p140: "Already the simple tactile qualities, such as soft and hard, and even more so rough and smooth, are not readily an instantaneous experience but require a series of changing sensations obtained… generally speaking by movement. Thus in their very constitution, a

discrete differences (information), it is digital in character. The world that it presents to us, however, appears to be analog. A fundamental dynamic of this virtuality is to integrate an apparent whole from digital information, by smoothing over, filling in, interpolating between, integrating over, and interpreting as irrelevant, the discreteness of neural events (as well as that of some fine-grained events in the external world).

Though qualia are products of complex neural events and reflect underlying structure, they gloss over and integrate such events and structure, masking their history or composition. Color or sound qualia synthesize many sources of information about structure in the environment to yield an appraisal that is useful to the organism. This appraisal is the finished product of a process; it bears little or no information about the process itself. Sensation implicitly and indirectly conveys information about properties such as wavelength of light, reflectance, sound frequency, etc.—which are explicitly measured by the scientific observer as quantities. But the individual wave fronts of light or sound are synthesized into a seemingly structureless emerging *quality*. This type of user-friendly presentation has been likened to the piloting interfaces of a modern aircraft, which allow attention to dwell in air space rather than on the detailed workings of the aircraft.

One may speculate that a global *picture* is built up essentially from primitive *responses* at a lower level, in the way that a digital image is built from pixels (MacLennan). Each "pixel" may represent a simple judgment (1 or 0), but the image that emerges with scale is an integration of that information, facilitating response on another level. The minute scale of the individual impulses (pixels) allows both their miniscule energy and their individual import to be subsumed into a larger synthesis from which the subject may be relatively detached—even though the "raw material is action: impacts, hustlings, clashes on a molecular level" (Jonas, p29).

The sense of complete detail in phenomenal experience is illusory. Mental images are poorly detailed compared to perceptual images, but even the latter are only relatively densely detailed. The illusion of limitless detail is underwritten by the fact that the senses can access further information, while memory or imagination cannot.[93] In both cases—of mental and of perceptual imagery—recognition consists in *recognizing that* something is the case. Reading a word, for example, is a matter of deciding that a particular configuration of marks on the page represents a word or phrase in a familiar language, which in turn represents a particular meaning.

Now, let's say that you, in your peripheral vision, notice a painting on the wall. Though you may have seen it every day, you see only *that* there is a "painting" there, but not its full detail, until the eyes are directed to center the fovea on the painting, and

---

synthesis on the part of the percipient is involved, extending over the time-span of the series and, by a short-term retention, unifying its elements into one impression."

[93] Suppose you close your eyes and behold a mental image of a printed page. Unless you have perfect eidetic memory, in this image you recognize *that* there are typed letters in lines, but not necessarily sufficient detail to read them. Perhaps distinct words are not made out, nor even the exact typeface. The mind's eye simply grasps *that* an imagined or recalled page contains printing. Now you open your eyes and look at a real printed page in a language you know. Obviously, you can access the actual words and sentences to read them. This you do by directing your (quite small) foveal vision at a line of print, one phrase at a time. However, in your peripheral vision, the words will not be clear, though you may anticipate them because of context.

attention is directed to search out more details within a limited visual area. Noticing a "detail," however, is much the same as noticing a painting, only on a finer scale. Seeing a detail is seeing that something is the case, just as you saw that it was a painting. The fact that one can identify indefinitely many such details (by moving the fovea about or by paying closer attention) distinguishes sensory perception from remembered imagery, imagination, or dreams.[94] But it is an illusion that perception reveals reality in all or even many details. In the case of vision, the illusion can be maintained because of rapid eye movements in the context of the relative constancy of the real physical scene (Crick and Koch 1998, p99). As Dennett (1991, p354-5) has pointed out, although a glance takes in with clarity only a very small portion of a repetitive pattern (on wallpaper, for example) yet it seems as though we uniformly see all the individual forms that are repeated.[95] I take this a step further to claim that it only seems as though we clearly see *even a single isolated form*.

These considerations suggest that perception *in general* is perception *that* rather than perception *of*. "Perception of" is not even a coherent notion, since it supposedly involves a straightforward presentation of what is, as though one passively surveys a scene. But perception entails categorizing what is on the retina. It is an act of the organism, the assertion of a proposition, a decision made about what is sensed. It may not be a *final* decision (as illustrated by the Necker cube and other ambivalent figures about which the brain cannot make up its mind). But, it is nevertheless a decision made with the tautological certainty of all definitions and decrees. Perception *needs* to be definite even when it is wrong. And it is only the organism's act of fiat that makes it so, and thereby makes consciousness possible.[96]

Filling in is a fundamental aspect of all perception, demonstrating the "digital" aspects of perception and the constructive capability of the brain to interpolate (Crick & Koch, 1992/2000, p15). Moreover, numerous laboratory experiments demonstrate visual completion effects of different sorts.[97] One of these experiments demonstrates a tactile version of filling in the visual blind spot.[98] Other experiments demonstrate various forms of spatial and temporal projection. Still others (among which the famous experiments of

---

[94] There are other distinguishing characteristics as well, such as consistency and changes in the visual field that conform with the idea of reality.

[95] Dennett's example is wallpaper by Andy Warhol, with a repeated image of Marilyn Monroe. Such an image is already literally iconic; because it is an artifact to begin with, it contains only limited detail. This is not the case with perception of a natural scene, which contains potentially unlimited detail.

[96] Frith & Metzinger: "There could be unconscious models… and they could certainly be characterized by a high degree of Bayes optimality. But only misrepresenting the probability of a hypothesis as 1.0 and simultaneously flagging it as a fact holding *now*… turns a possibility (or a likelihood) into a reality. This is what makes the zombie conscious."

[97] For a "taxonomy of perceptual completion phenomena," see: Pessoa et al (1998).

[98] O'Regan & Noë, p958: "For practical reasons the battery of 400 vibrators mounted on the observer's back consisted of two ramps of 200 vibrators, one on each side of the observer's backbone. A large gap was therefore present in the tactile representation of the visual field….[but] no gap was apparent in observers' perceived visual field. This [is a] tactile analog of what might incorrectly be called 'filling-in' of the retinal blind spot…"

Stratton,[99] oft repeated in variations) show the adaptability of the nervous system to restore perception that corresponds to functional behavior within an environment. Some lead to quite bizarre experiences, such as the disorientation first experienced by subjects wearing lenses that invert or reverse the optical input to the eyes;[100] or the strange feeling of phantom limb that can be induced in the "rubber hand" experiment. Apparent motion effects are well known in modern culture through films, illuminated signs, etc. Usually artificially produced, such effects are considered illusory when they involve (perceptually) jumping to a *false* conclusion. But the general lesson is that *all* perception is naturally a matter of jumping to conclusions, whether warranted or not. Phenomenality might then be characterized as a useful illusion realistically guided by sensory input.

8.3 Qualia and introspection

The problems posed by qualia[101] seem to have aroused more passionate discussion than any others in the philosophy of mind (Armstrong, p121). This is scarcely surprising, since qualia are what a theory of phenomenality should explain (Crane, p171). The term 'qualia' (singular 'quale') was first used in its modern sense by C. S. Peirce in 1866. But explicit focus on the role of qualia in the mind-body problem only began in 1958 with Herbert Feigl's *The Mental and the Physical* (Crane, p177-81).

Qualia are considered properties of what is "given" in experience. This is partly what makes the notion problematic—like the disgraced notion of 'sense-data'. For, what is "given" is not "raw material" (input) for processing but rather a "finished product" (output) for presentation to an executive function. Moreover, qualia are not *properties of* experience (for, as we have seen, properties are propositional facts), but simply the experience itself. Yet, since perception seems propositional insofar as it notes features, it is sometimes (absurdly) argued that qualia representing this propositional information are epiphenomenal, or even that they don't actually exist.[102] Quite the contrary, I have argued

---

[99] Here is Stratton's (1896) own report: "As to the relation of the visual field to the observer, the feeling that the field was upside down remained in general throughout the experiment. At times, however, there were peculiar variations in this feeling according to the mental attitude of the observer toward the present scene. If the attention was directed mainly inward, and things were viewed only in indirect attention, they seemed clearly to be inverted. But when, on the other hand, full attention was given to the outer objects, these frequently seemed to be in normal position, and whatever there was of abnormality seemed to lie in myself, as if head and shoulders were inverted and I were viewing objects from that position, as boys sometimes do from between their legs."

[100] While not all creatures can achieve this adaptation, apparently even some robots can (Boden 2006, p1325-27, citing work of Ezequiel Di Paolo).

[101] According to Wikipedia, qualia are "individual instances of subjective, conscious experience." In other words, they are the individualized contents of phenomenal experience. I have often avoided the term in favor of the older 'quality', which dates to usage by the early natural philosophers. They distinguished secondary qualities from the primary ones that were supposed to be quantitatively expressed. 'Qualia' has since been associated with the secondary qualities, in contrast to supposedly objective properties of the external world.

[102] Dennett (1991, p372) does not clarify matters by denying that qualia exist while agreeing that they *seem* to. The question of whether qualia "exist" bears some resemblance to the question of whether quanta really exist. Both are ambiguous as to whether they should be thought of as

that qualia themselves represent what can be taken as propositional information—in the form most useful to the executive function, which is then able to exert downward causation leading to action.[103]

The issues surrounding qualia are understandably confusing, since one does not normally and naturally see such things as "color patches" or have "raw feels." One cannot readily factor out sensation from perceptual interpretation (Jack & Shallice, sec5; Loar, p9).[104] Qualia are not objects of perception, but integral aspects of it.[105] When sensation seems to be in itself an object of attention, isolated from perception, it is by virtue of a special attitude or way of looking. With special effort, for example, it is possible to "deconstruct" features of visual experience in order to attend to color patches instead of colorful objects. Artists routinely do this, and we are culturally accustomed to the idea of pure color through the long use of paints and dyes. (Similarly, meditators are able to dissociate from the discomfort of body sensations precisely by attending to them as such.) Color patches and raw feels are not the input to cognition but the result of special cognitive acts. In experimental situations, qualia are literally artifacts of the experimental set-up. For instance, one does not normally encounter monochromatic color patches in isolation from natural objects and surfaces, yet such are routinely used in experimental psychology. So, the very *notion* of qualia is a byproduct of the possibility of introspection, on the one hand, but also of literally artificial environments, on the other.

Part of the confusion surrounding qualia is that the term can be understood to mean either objects of introspection or simply "what it is like" to be having a phenomenal experience. (There is something it's like to see red whether or not one is self-consciously introspecting.) Introspection in turn can be understood variously as a special way of looking at experience (bracketing), as a form of self-observation (Sloman & Chrisley, 2003, sec8.1), as self-reporting, or as the information processing system "talking to itself about itself" (Boden 2006, p1238).

---

substantial and individualized things or as mere aspects of a continuum. Neither indicates a stable entity over time, but fleeting appearances better serving as evidence for physical properties than as entities in their own right. Perhaps alluding to a comparison with quantum uncertainty, Dennett (1991, p138) shrewdly observes that "…there are no fixed facts about the stream of consciousness independent of particular probes."

[103] Cf. Boden (2006, p1421): "Sloman discussed the causal relations within a wide variety of virtual machines, and between the virtual and the physical machine…One of his conclusions was that a virtual process can properly be said to cause a physical one, so that qualia—which he analyses in computational terms… aren't epiphenomenal, but really do cause changes in the brain."

[104] On the dubious notion of sensation as primitive input, cf. Edwards & Wilcox, sec3: "There is no evidence whatsoever for a primitive experience of meaningless points of light or colored shapes from which we infer our perceptions." That is, such a phenomenological concept of sensation is incoherent; there is need for a concept of input to the nervous system, a definition of sensation in physical terms. King (2007, p188 points out that, historically and etymologically, "the word 'sensation' is not an ordinary English word. It is rather a bit of philosophical jargon, a technical term specifically introduced to talk about phenomenal content independent of its (external) cause (if any)."

[105] Cf. Humphrey (2000a): "Sensations are no more the objects of sensing than, say, volitions are the objects of willing or intentions the objects of intending."

To introspect, moreover, is not literally to "look within"—i.e., to regard experience as presented in an inner theater—what U.T. Place called the "phenomenological fallacy" (Boden, 2006, p1344). There is no need for an internal audience or theater to access information, because it is not images but data sets that are accessed, like those in a computer memory from which a video display can be constructed.[106] Yet, because introspection is a conscious act it does imply a conscious agent, which is not a witness *within* the person but the conscious person herself. There are sensory referents for this act. Just as there are proprioceptive referents for the sense of willing and moving a limb or directing the eyes, there is something it is like to be introspecting. Deliberate scrutiny of something in the visual field, for example, is a different experience from deliberately softening the gaze, or from just plain looking.

Because qualia refer to the organism's co-determined *interactions* with physical reality, Jackson's (1986) "Mary" can know everything known about physical processes or properties and not about corresponding qualia. Phenomenal experience is not deducible from physical facts alone because consciousness is not a product of the object alone (but of the subject interacting with the object), and because physical facts are designed to exclude qualia in the first place. Our naively realist orientation is toward static properties of the world more than dynamic interactions of world and organism. Indeed, Robbins (p157-8) has pointed out that qualia seem predominantly static rather than presenting aspects of motion or change.

We do not normally experience sensation without some perceptual interpretation.[107] Yet—pathologically—there *can* be sensation without perception (e.g., visual agnosia, the inability to recognize objects seen, and some kinds of neuralgia); and there can be perception without sensation (e.g., "blindsight", the ability to recognize things when they are not owing to cortical damage, consciously seen). Sensation may be considered the "evidence" on which perception is based. Blindsight subjects feel they are just guessing at questions about what they do not consciously see (and for which they thus think they have no sensory evidence) but which they manage, at least statistically, to recognize. They are unable to take responsibility for perceptual acts that do not seem to be acts of the conscious self, justified by conscious sensation. Agnosia, on the other hand, is the inability to make anything of the evidence one does have.

Functionalism is deemed powerless to give an account of conscious experience when qualia are *assumed* to be without function—that is, to play no role in the brain's input-output transformation of information (Revonsuo, p38). However, if one assumes rather that qualia contain not only information about the world, but also information concerning the input-output relations (e.g., relations experienced subjectively as affect), then qualia are definitely functional and must be included in the information processing cycle that is traditionally conceived in behaviorist terms. They represent *to the organism* information that the observer conceives to play a functional role.

---

[106] Lehar, sec6.1: "If a picture in the head required a homunculus to view it, then the same argument would hold for any other form of information in the brain, which would also require a homunculus to read or interpret that information. But, in fact, any information encoded in the brain needs only to be available to other internal processes rather than to a miniature copy of the whole brain." Ironically, it was Descartes who first realized this.

[107] Pain has been used as a counterexample, but its built-in affect *constitutes* an interpretation: as tissue damage, with appropriate protective behavior.

Qualia provide information about the world—but not only about the world, nor in explicitly propositional form. Redness, for example, bears information concerning light of certain wavelengths. But it also indicates information about the channel of information, which is the medium through which the subject knows the world using only the equipment provided naturally at birth. While wavelength can be measured in various ways by means of other equipment, the *quality* of redness cannot be known by any means except direct experience through the subject's visual channel. Other equipment, such as a spectrometer, provides a different access to knowledge about the world, effectively through a different sensory channel in a different cognitive domain. If such equipment were actually to become part of a conscious cyber-organism, we might expect that the channel of information it provides would have its own proper "quality," since this is no more than an extrapolation from the experiential evidence provided by existing prosthetic substitutions.

8.4 Absent and inverted qualia

Descartes and La Mettrie imagined *l'homme machine* as a human body without a soul—able to respond to stimulation as people do but without consciousness. This prefigured the modern philosophical notion of the *zombie*, a favorite thought experiment used to argue concerning the functionality of consciousness. If consciousness is merely epiphenomenal, then an ersatz version of a human being could be structurally and behaviorally identical without being conscious.

Some have argued that, because zombies are "conceivable," they are "metaphysically possible" (Chalmers, 2002, sec 3). Others have argued that they are not even conceivable (Tye, p153). Conceivability, however, seems more a psychological than a logical question, and more a function (or dysfunction) of language than of (visual) imagination. Invented possibilities can be stated in the same grammatical form as ordinary facts, and thus pass as "possible" on the same footing, whether or not they are self-consistent notions. Such conceptual acts may gloss over the details required to render the possible actual. While one can conceive of a system that does "all the same" information processing as humans, we may not (currently, or perhaps ever) be able to imagine it in sufficient detail to recognize the functionality of those processes uniquely associated with phenomenality (Chalmers, 2002, sec 7). Moreover, if 'zombie' describes something structurally and functionally identical to a normal person, then, by the identity of indiscernibles, one should expect zombies to be as conscious as one expects people to be. On the other hand, biologically normal humans can apparently function quite well without full consciousness (sleepwalking, driving while daydreaming, etc.), so one should expect the same range of possible behavior of their ersatz version.

Of two "functionally identical" systems, both should have phenomenality if one of them does. However, that is an empty claim if one of them is a *natural* system, since there is no way to establish perfect identity of a system that is not well-defined. Exhaustively establishing such identity is at best possible between *artifacts*—because each is finitely specifiable, whereas natural things are not (Bruiger, 2016). The absent qualia argument holds that complete physical description does not entail phenomenal consciousness, because such a description would be equally consistent with no consciousness at all. Hence, on that argument, physicalism is at least incomplete, if not false. But it is a moot point because no complete physical description *can* be given of any natural system, let alone the brain.

Can someone possessing only scientific information about perception know "what it is like" to be having the experience that corresponds to that information? A totally sequestered scientist in a black and white lab, when finally released to experience color for the first time (Jackson 1986), does not gain knowledge of the *world* from that experience that she could not gain otherwise. What she does gain is knowledge of her own cognitive system in a new interaction with the world, involving her body's instrumentation and her brain's "language of the senses," when exposed to colored objects. On the other hand, the very point of scientific instrumentation is to free the observer from such bodily idiosyncrasies. Natural perception itself attempts in various ways (such as object constancy) to achieve this freedom, which the scientific method extends (Mausfeld, sec1). The problem is that it comes at the price of then seeming alienated from the qualitative aspects of embodied experience.

If epiphenomenalism were true, then we could not be certain, by behavioral signs alone, that another creature or person experiences pain. However, if pain is functional, it is because it signifies a valuation concerning the state of one's wellbeing. It is possible for an outside observer, such as a doctor, to make an assessment of damage apart from the question of the patient's pain. If we assume that the organism itself tracks and evaluates its own state in experience, then the possibility of pain must also be assumed.

Related considerations apply to the notion of the inverted spectrum—or, more generally, the idea that qualia could be arbitrarily or systematically interchanged among different human subjects in a behaviorally undetectable way.[108] Of course, there are differences in human sensitivities; some cannot distinguish between certain colors, for example. But the inverted spectrum conundrum is not about behaviorally observable sensitivities. While it is not a *scientific* question (since, by supposition, it cannot be detected behaviorally), yet on logical grounds there can be no inverted spectrum. This is because the particular subjective experience of a given color *means* something determinate in "the language of the senses," just as particular words mean something specific in a given natural language. It would not be the same language if the referents of words were systematically interchanged. If humans inherit the same neural "language" on account of common physiology, then the real-world referents of qualia must be assumed to hold, more or less, over individuals of the species.

8.5 Virtuality

Simulation always involves a loss of real detail, since the simulation is a finite program whereas natural reality is indefinitely detailed. Perhaps one reason why virtual reality is engaging—and its coarse-graining acceptable as entertainment—is because sensory experience itself is normally so impressionistic. When attending to a visual scene, for

---

[108] Descartes himself proposed this: "God could have made the nature of man such that this particular motion generated an entirely different kind of sensation in the mind" (quoted in Cottingham, p141).

example, it is the overall impression and the feeling it evokes that constitute the experience, not a comprehensive survey of all detail. It is an illusion that we fully and uniformly *see* all that is before the eyes; rather, what we "see" is what we notice—in some cases simply *that* there is detail, without putting a finer point on it. To the extent that a virtual reality can present fine enough detail to give the requisite illusion, it can pass as a substitute for sensory experience. Similarly, an artist's crude brushstrokes can produce what appears from a distance to be a highly detailed scene characterized by continuity—an effect that gave rise to the term 'impressionism.'[109]

Just as computation has served fruitfully as a metaphor for the workings of the brain underlying behavior, so virtual reality provides a metaphor to understand how the brain, encapsulated in the skull, can produce the spectacle we call phenomenality, which includes experience of the body.[110] However, one would miss the point to think of "natural" virtual reality as a sort of "goggles" that the brain puts on to have an experience, as the human user does. Or, to imagine the vista presented by the VR as though dancing inside the brain like a holographic image or a miniature replica of the external world (Lehar, sec10): a Cartesian theater. The very point of the VR metaphor is that the brain *is* the device that generates the virtual reality—not for some spectator but for its own purposes and benefit, which is not entertainment but survival. The "simulation" cannot be compared to the reality it simulates, because the brain cannot get outside the skull to make the comparison. So, the question of verisimilitude can only be measured (after the fact) by the evolutionary success it facilitates. Nothing is accessible outside the simulation itself to set a standard of comparison by which to decide the question, except the truths of science, which are themselves part of the virtual reality and which must ultimately also be measured by evolutionary success.

It is not shapes and colors falling upon the retina that we experience, but the appearance of a real world around us. This appearance is largely an improvisation based on relevant sensory information, continually updated. In other words: a virtual reality based on real-time reality. By its very nature, the *fiat* I have described as a filling-in or completion process is not a matter of causal pathways but of *logical* connections. No strict isomorphism between the virtual reality and the physical system on which it runs need be assumed, because the relationship is specified by the system itself: its connections are topological rather than topographic, with whatever meanings the system has established them to mean.

9.0 Artificial cognition

The computer metaphor serves to help understand cognition from an engineering point of view. A scientific model is a conceptual version of a possible engineering project, the most convincing test of which would be to actually construct it. A model of an intentional system is "realistic" to the degree a physical realization of it can be constructed that "replicates" the behavior of the system. This strategy brings into play the advantage

[109] It seems we are more gratified by success at creating a simulacrum than we are amazed by how impoverished a depiction our ordinary perception usually is.
[110] We no more "directly" experience the body than other parts of the external world (Revonsuo, p194). The body, as experienced, is a part of the virtual reality.

traditionally known as *maker's knowledge* (after Vico), which is the insider's knowledge the creator has of the creation. Epistemically, natural systems are ambiguous, while human artifacts are products of human specification and therefore well defined in principle. We can be assured of understanding a system we have designed ourselves; to the degree it does what its natural counterpart does, we can be relatively assured of understanding the latter too. The hitch lies in the "degree," of course, which is relative and can be deceptive. Hence, the notion of replication is ill defined. A machine that pitches baseballs for batting practice does "the same thing" as a human pitcher, but without the same internal workings. It would be a mistake to think that the machine adequately models the person or even the action of "pitching."

In any case, artificial systems that *design themselves* would not offer this insider's knowledge. An autopoietic machine (self-defining, self-organizing, self-maintaining) would be an *organism* even if made of inorganic materials. It would not be a product of human definition, would not be any more (or less) readily comprehensible than other natural things. It would be subject to human control only in the ways that natural things are, if at all. Yet, within these limits, an autopoietic machine would convincingly demonstrate that life and intelligence depend more on organization than materials, supporting functionalism. In regard to the study of mind, the ability to set in motion such an eventuality would provide "the chance to witness, with our own eyes, how intentional capacities can arise in a 'merely' physical mechanism" (Smith, p18).

The other side of the zombie coin is the intuition that machines could be conscious. A challenge of the hard problem is to justify that intuition by showing how phenomenality could be artificially achieved. Certainly, the ultimate power of the computational metaphor in this regard rests on the promise that machines could convince us that they have phenomenal experience. The computational metaphor provides an engineering strategy to bridge the explanatory gap. If it could be cashed out, it would offer the satisfaction that the brain's natural everyday bridging of the gap has been recapitulated artificially, even if we may not know exactly how.

The trouble with that promise has become apparent with large language models (chatbots), which give the appearance of conversation with a fellow human being. It is through language that we humans convince one another that we are conscious, that we feel and have phenomenal experience. Since LLMs are by definition masters at language, they could be similarly convincing, simply by saying the right things. A better guide than any version of the Turing Test would be to know, on grounds independent of language, that the communicating system is an autopoietic agent.

The notion of creating life, intelligence, and even consciousness from scratch may be motivated by the desire to imitate and control nature, to play God, and to exceed natural human limits. Artificial intelligence is not just a project to make something functionally resemble a natural counterpart, but ultimately to produce the real thing. In terms of phenomenality, the goal would be to create a system that actually feels and consciously experiences, not just one that behaves as though it does, says it does, or is programmed to convince us it does.

10.0 Summary conclusions

I have surveyed some of the reasons why the hard problem is hard, and have assembled threads of a loose argument that a solution to it requires taking the point of view of the organism and its subagents; I argued further that doing so need not pose any regression of observers or a Cartesian theater. The gist of the argument is that consciousness is not something that *happens*, through passive efficient causation in the brain, but something the brain (or the organism) intentionally *does*. This action I have called *fiat*, which is the organism's counterpart to the human or divine ability to simply declare things into being. To be conscious, a creature (whether natural or artificial) must be an autopoietic system, with motivations that arise through an embodied evolutionary history. More particularly, consciousness is a virtual reality at the service of an executive agent that monitors the organism's processes in relation to a putative external world. Phenomenality and consciousness are therefore not epiphenomenal but serve a definite function, distinct from unconscious processing and capable of downward causation.

Does this approach close the explanatory gap? Perhaps we have come as close as possible to a scientific explanation of consciousness, but still with no cigar. For one thing, scientific explanation has limited its scope by eliminating subjectivity from its discourse, while subjectivity is the very essence of consciousness. Science has also banned agency and teleology in favor of a view of matter in terms of efficient causes. A third-person *observer's* stance has become the basis of scientific description, which is fundamentally incompatible with understanding the first-person point of view of a conscious organism. In order to enable a scientific explanation of consciousness, science might have to redefine itself in several ways. First of all, it could examine its own biases, fundamental dogmas, and motivations. It could forfeit the pretense to objectivity, grasping that its epistemic situation is no different from that of the brain it studies, in terms of whose activities it hopes to explain consciousness. It could broaden its view of matter as essentially passive. That view has never fit organisms and may no longer fit a widening range of self-organizing processes in the non-living world as well. It certainly does not fit the concept of agency, as introduced here, which I believe essential to understand consciousness.

Yet, even granting these modifications, it is possible we may never find *any* explanation of consciousness satisfying. Perhaps no approach can tell us why there is such a thig as experience in the first place (Chalmers,1995, p20). We may never be satisfied that the ineffable, self-luminous quality of phenomenal experience is accounted for by any merely intellectual maneuver. We may remain unconvinced by any reduction of first-person experience to third-person descriptions. This impasse may prove to be an inescapable aspect of our life as self-conscious beings with thought based on fully grammatical language. The explanatory gap between thought and experience may be too wide, and the idea of engineering a conscious robot may be too problematic, if not paradoxical. For, to constitute an organism, an artificial creature would have to *self*-engineer; it would have to be a product of some artificial equivalent of natural selection.

Perhaps we will always wonder how it is that we happen to be *conscious,* in the same way we may always wonder how it happens that there is *anything at all*. Yet, as Chalmers points out : "But this is the same for any fundamental theory. Nothing in physics tells us why there is matter in the first place, but we do not count this against theories of matter." Despite the advances of science, we may never have an answer to the mystery of why anything exists at all. Similarly, we may never have an answer to the mystery of how there is consciousness of it. It is only through our astonishing reflexivity that the two questions can even be posed.

The scientific interpretation of nature parallels the epistemic challenge of cognition generally. So, the question of the nature of consciousness contains the question of the nature of science—and, therefore, the question of the ultimate nature of natural reality. If the brain seems a black box to investigators outside of it, the world outside the skull is equally a black box to the brains of scientists. The brain can only compare its own inputs and outputs to infer whatever is going on "out there." So much is true for the scientist as well, who can but compare instrumental settings and readings. There is an important difference, however. In the case of natural perception, we come literally to experience our cognitive processing as the real external world itself. So to speak, we live in the model as in a virtual reality, as though access to the model granted transparent access to the world it mediates. This "illusion" seems a justified strategy of evolutionary history, without which we would not be here. Scientific models also come easily to be taken for what they model, but with the important difference that wrong scientific theories are not usually fatal. On the other hand, the fact that a wrong-headed theory hasn't killed us yet doesn't make it true.

REFERENCES

Arico, Adam; Fiala, Brian; Goldberg, Robert F.; & Nichols, Shaun (2011) "The Folk Psychology of Consciousness" Mind & Language, Vol. 26, No. 3 June 2011

Armstrong, D. M. (1999) *The Mind-Body Problem: an opinionated introduction* Westview (Perseus)

Ayer, A. J. (1946) *Language, Truth, and Logic* Dover reprint

Bateson, Gregory (1972) *Steps to An Ecology of Mind* Ballantine

Beaton, Michael J. S. (2009) "An Analysis of Qualitative Feel as the Introspectable Subjective Aspect of a Space of Reasons" (Doctoral thesis)

Beaton, Michael (2016) "Crossing the Explanatory Gap by Legwork, not by Fiat" *Constructivist Foundations* vol 11, no. 2, March 2016

Beaudoin, Luc (1994) "GOAL PROCESSING IN AUTONOMOUS AGENTS" (PhD thesis, p5) http://www.sfu.ca/~lpb/pubs.html

Bennett, Bruce M., Hoffman, Donald D., Prakash, Chetan (1989) *Observer Mechanics: A formal theory of perception* Academic Press.

Boden, Margaret A. (1970) "Intentionality and Physical Systems" *Phil. Of Science* Vol 37, No. 2, June 1970

Boden, Margaret A. (2006) *Mind as Machine: A History of Cognitive Science*, Volumes 1&2, Oxford University Press

Boltuc, Nicholas and Boltuc, Piotr (2007) “Replication of the Hard Problem of Consciousness in AI and Bio-AI: An Early Conceptual Framework” preprint for Anthony Chella & Ricardo Manzotti (eds.), *AI and Consciousness: Theoretical Foundations and Current Approaches*. Aaai Press, Merlo Park, Ca

Botvinick, M. and Cohen, J. (1998) *Nature* 391, 756 (19 February 1998), cited in Metzinger 2010, p40

Bruiger, Dan (2014) “An Argument for a Second-Order Cosmology” www.academia.edu/7963044/An_Argument_for_a_Second-Order_Cosmology

Bruiger, Dan (2016) *The Found and the Made: science, reason, and the reality of nature* Transaction Pubishers

Bruiger, Dan (2017) “Causes, Goals, and Reasons: Clarifying the Meanings of Teleology” Foundational Questions Institute, http://fqxi.org/community/forum/topic/2725

Bruiger, Dan (2024) “Understanding Consciousness by Analogy” *Journal of Mind and Behavior* Vol. 45, Nos. 3 and 4, pp301-312.

Chalmers, David J. (1995) “Facing Up to the Problem of Consciousness” http://consc.net/papers/facing.html, (postprint from *Journal of Consciousness Studies* 2(3):200-19

Chalmers, David J. (2002) “Consciousness and its Place in Nature” published in D. Chalmers (ed.) *Philosophy of Mind: Classical and Contemporary Readings* Oxford UP

Cottingham, John (2000) “Intentionality of Phenomenology? Descartes and the Objects of Thought” in *History of the Mind–Body Problem*, edited by Tim Crane and Sarah Patterson, Routledge

Crane, Tim (2000) “The Origins of Qualia” in *History of the Mind–Body Problem* Edited by Tim Crane and Sarah Patterson Routledge

Crane, Tim and Patterson, Sarah (2000) “Introduction” in *History of the Mind–Body Problem,* edited by Tim Crane and Sarah Patterson, Routledge

Crick, Francis and Koch, Christof (1992/2002) “The Problem of Consciousness” *Scientific American* 2002, updated from September 1992

Crick, Francis and Koch, Christof (1998) “Consciousness and Neuroscience” *Cerebral Cortex* Mar 1998;8:97–107

Crick, Francis and Kock, Christof (2003) "A Framework for Consciousness" *Nature* http://www.klab.caltech.edu/koch/crick-koch-03.pdf

Dehaene, Stanislas (1997) "What Are Numbers, Really? A Cerebral Basis For Number Sense", reprinted in *Edge* March 19, 2016

Dennett, Daniel C. (1978) *Brainstorms* MIT Press

Dennett, D.C. (1981) "Intentional Systems" in *Mind Design*, John Haugeland, ed., MIT Press

Dennett, D.C. (1987/1998) *The Intentional Stance* MIT Press

Dennett, D.C. (1991) *Consciousness Explained* Little Brown & Co

Dretske, Fred (1994) *Naturalizing the Mind* MIT Press

Dreyfus, Hubert L. (2007) "Why Heideggerian AI Failed and How Fixing it Would Require Making it More Heideggerian", Sec III, preprint for *Philosophical Psychology* 171(18):1137-1160

Eady, Trent (2015) "Climbing to Consciousness: The Mind-Body Problem and the Computational Order" *Res Cogitans* (2015) 6:14-20

Edwards, David A. and Wilcox, Stephen (1980) "Unity, Disunity and Pluralism in Science" arxiv.org/pdf/1110.6545

Eisemann, C. H.; Jorgensen, W. K.; Merritt, D. J.; Rice, M. J.; Cribb, B. W.; Webb; P. D.; and Zalucki, M. P. (1984) "Do Insects Feel Pain? – A biological view" Experientia 40

Ellis, Ralph (2000)"Efferent Brain Processes and the Enactive Approach to Consciousness" *Journal of Consciousness Studies* 7, no. 4, 2000

Faichney, Robin (2013) "Mind, Matter, Meaning and Information" *TripleC* 11(1): 36-45

Feigl, Herbert (1958) "The Mental and the Physical" *Minnesota Studies in the Philosophy of Science* 2:370-497

Frith, Chris D. and Metzinger, Thomas K. (2013) "What's the Use of Consciousness?" https://www.blogs.uni-mainz.de/fb05philosophie/files/2013/04/Frith_Metzinger_Regret_2016_penultimate.pdf)

Grahek, Nikola (2001) *Feeling Pain and Being in Pain* Oldenburg: Hanse Institute for Advanced Study, p150, quoted in Tim Crane "Cosmic Hermeneutics vs. Emergence: The Challenge of the Explanatory Gap" http://philpapers.org/archive/CRACHV-2.pdf

Gray, Jeffrey A. (1998) “Creeping up on the Hard Question of Consciousness” in Hameroff et al

Grush, Rick (2003) “In Defense of Some ‘Cartesian’ Assumptions Concerning the Brain and Its Operation” *Biology and Philosophy* **18:** 53–93, 2003

Hameroff, Stuart R. et al, eds. (1998) *Toward a Science of Consciousness* MIT Press

Hameroff, S. and Penrose, R. (2014) “Consciousness in the universe: a review of the 'Orch OR' theory” *Phys Life Rev* 2014 Mar;11(1):39-78

Heidelberger, Michael (2003) “The Mind-Body Problem in the Origin of Logical Empiricism: Herbert Feigl and Psychophysical Parallelism” in *Logical Empiricism: Historical and Contemporary Perspectives*, Paolo Parrini, Wesley C. Salmon, and Merrilee H. Salmon (eds.). University of Pittsburgh Press

Helmholtz, Herman von (1867) *Handbuch der Physiologischen Optik*, p.443-46, quoted in Mausfeld 2002, appendix p39

Hirstein, W. (2011) “The Contribution of Prefrontal Executive Processes to Creating a Sense of Self” In: *Brain, Mind and Consciousness: An International, Interdisciplinary Perspective* (A.R. Singh and S.A. Singh eds.), *MSM,* 9(1)

Hoffman, Donald D. (2008) “Conscious Realism and the Mind-Body Problem” *Mind & Matter* Vol. 6(1), pp. 87–121

Hoffman, Donald D. (2009) “The Interface Theory of Perception”, sec 1.4 (A chapter for the book "Object Categorization: Computer and Human Vision Perspectives," edited by Sven Dickinson, Michael Tarr, Ales Leonardis and Bernt Schiele. Cambridge University Press, 2009, pages 148-265)

Hoffman, Donald. D. (2011) “The Construction of Visual Reality”, preprint for the book "Hallucination: Theory and Practice", edited by J.D. Blom and I. Sommer. Springer Verlag

Hoffman, Donald D., and Prakash, Chetan (2014) “Objects of consciousness” *Frontiers in Psychology* June 2014 | Volume 5 | Article 577

Horst, Steven (1999) “Evolutionary Explanation and the Hard Problem of Consciousness” *Journal of Consciousness Studies* 6, No 1, 1999

Hohwy, J. (2013) *The Predictive Mind.* Oxford UP.

Humphrey, Nicholas (1976) “The Colour Currency of Nature” First published in *Colour for Architecture*, ed. Tom Porter and Byron Mikellides, pp. 95-98, Studio-Vista, London

Humphrey, Nicholas (1992) *A History of the Mind: Evolution and the birth of consciousness*. Simon & Schuster.

Humphrey, Nicholas (2000a) “THE PRIVATIZATION OF SENSATION” In *The Evolution of Cognition*, ed. L. Huber and C. Heyes, pp. 241-252, MIT Press

N. Humphrey (2000b) *How to Solve the Mind-Body Problem.* Imprint Academic.

Huxley, Thomas (1874) “On the Hypothesis that Animals Are Automata, and Its History” *Collected Essays I*

Huxley, T. H. (1886) *Lessons in Elementary Physiology*. London: Macmillan

Iglowitz, Jerome (2010) *Virtual Reality: consciousness* really *explained*, third ed, Jerrysplace Publishing

Jack, Anthony I. (2013) “A scientific case for conceptual dualism: The problem of consciousness and the opposing domains hypothesis.” preprint

Jack, Anthony I. and Shallice, Tim (2001) “INTROSPECTIVE PHYSICALISM AS AN APPROACH TO THE SCIENCE OF CONSCIOUSNESS” preprint for *Cognition*, Vol 79, Issue 1-2, p161-196

Jack, Anthony I.; Robbins, Philip; & Roepstorff, Andreas (2007) “The genuine problem of consciousness” *Philpapers*

Jackson, Frank (1982) “Epiphenomenal Qualia” Reprinted from *The Philosophical Quarterly*, 32, No. 127. (Apr., 1982), pp. 127-136

Jackson, Frank (1986) "What Mary Didn't Know" *The Journal of Philosophy.* **83** (5): 291–295

Jonas, Hans (1966/2001) *The Phenomenon of Life: toward a philosophical biology* Northwest University Press

Kandel, Eric (2012) *The Age of Insight: the quest to understand the unconscious in art, mind, and brain.* Random House.

Kenny, Vincent (2009) “There’s Nothing Like the Real Thing: Revisiting the Need for a Third-Order Cybernetics.” *Constructivist Foundations* 4:103.

King, Peter (2007) “Why Isn’t the Mind-body Problem Medieval?” in H. Lagerlund (ed.), *Forming the Mind*: *Essays on the Internal Senses and the Mind/Body Problem from Avicenna to the Medical Enlightenment*, Springer 2007, 187–205

Kouider, Sid; de Gardelle, Vincent; Sackur, Jerome; & Emmanuel Dupoux "How rich is consciousness? The partial awareness hypothesis" (2010) *Trends in Cognitive Sciences* Vol.14 No.7

Kriegel, Uriah (2012) "The Phenomenal Intentionality Research Program" In U. Kriegel (ed.), *Phenomenal Intentionality*. Oxford University Press (2013)

Kuipers, Benjamin (2007) "Sneaking Up On the Hard Problem of Consciousness" https://www.aaai.org/Papers/Symposia/Fall/2007/FS-07-01/FS07-01-015.pdf

Lamme, Victor (2015) "The Crack of Dawn: Perceptual Functions and Neural Mechanisms that Mark the Transition from Unconscious Processing to Conscious Vision". In T. Metzinger & J.M. Windt (Eds). Open MIND:22(T)

Lehar, Steven (2003) "Gestalt isomorphism and the primacy of subjective conscious experience: A Gestalt Bubble model" *Behavioral and Brain Sciences* 26, p376, (discussing Pessoa et al 1998)

Levine, Joseph (1983) "Materialism and Qualia: the explanatory gap" *Pacific Philosophical Quarterly* 64 (October): 354-61

Libet, Benjamin; Wright Jr., Elwood W.; Feinstein, Bertram; Pearl, Dennis K. (1979) "Subjective Referral of the Timing for a Conscious Sensory Experience - A Functional Role for the Somatosensory Specific Projection System in Man". *Brain* **102**: 193–224

Loar, Brian (1990) "Phenomenal States" *Philosophical Perspectives* 4:81-108

Lockwood, Michael (1998) "The Enigma of Sentience" in Hameroff et al (1998)

Long, Douglas C. (2010) "Why Life is Necessary for Mind: The Significance of Animate Behavior" (preprint for chapter in *Self, Language, and World: Problems from Kant, Sellars, and Rosenberg*. J. O'Shea & E. Rubenstein, eds (Ridgeview Publishing Company), 61-88)

Loorits, Kristjan (2014) "Structural qualia: a solution to the hard problem of consciousness" *Front Psychol* 5: 237

Lycan, William G. (2008) "Phenomenal Intentionalities", p1, preprint for *American Philosophical Quarterly* Vol. 45, No. 3 (July 2008) pp. 233-252

MacLennan, B. J. (2005) "Consciousness in Robots: The Hard Problem and Some Less Hard Problems", Sec 2B, preprint of ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication

Mark, Justin T.; Marion, Brian B.; and Donald D. Hoffman (2010) "Natural selection and veridical perceptions" *Journal of Theoretical Biology* 266, 504–515

Martin, M.G.F. (2000) “Beyond dispute: sense-data, intentionality and the mind–body problem” in *History of the Mind–Body Problem* Edited by Tim Crane and Sarah Patterson Routledge

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

Mausfeld, Rainer (2002) “The Physicalistic Trap in Perception Theory” in: D. Heyer & R. Mausfeld (eds.) *Perception and the Physical World* Wiley

Maxwell, Nicholas (2000) “The Mind-Body Problem and Explanatory Dualism” *Philosophy* Vol. 75, No. 291 pp. 49-71

McCulloch, Warren S. and Pitts, Walter (1990) “A Logical Calculus of the Ideas Immanent in Nervous Activity” *Bulletin of Mathematical Biology* Vol 52, No 1/2

McGinn, Colin (1989) “Can We Solve the Mind-body Problem?” *Mind* New Series vol 98 no 391, pp 349-366

Metzinger, Thomas (2005) “Précis: Being No One” *Psyche* 2005: Vol 10 issue 1

Metzinger, T. (2009) *The Ego Tunnel: the science of mind and the myth of the self* Basic Books

Metzinger, T. (2010) “The Self-Model Theory of Subjectivity: A Brief Summary with Examples” *Humana.Mente* – Issue 14 – July 2010

Metzinger, T. (2013) “The myth of cognitive agency: subpersonal thinking as a cyclically recurring loss of mental autonomy” *Frontiers in Psychology* December 2013 | Volume 4 | Article 93

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, pp783–792

Newton, Isaac (1730) *Optics* Book One, Part Two, Proposition 2

Ney, Alyssa (2016) “Microphysical Causation and the Case for Physicalism” preprint for *Analytic Philosophy* Vol 57 No 2 June 2016

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

O'Regan, J. Kevin and Noë, Alva (2001) "A sensorimotor account of vision and visual consciousness" *Behavioral and Brain Sciences* (2001) 24, 939–1031]

Pessoa, Luiz; Thomson, Evan; & Noë, Alva (1998) "Finding out about filling-in: A guide to perceptual completion for visual science and the philosophy of perception" *Behavioral and Brain Sciences* 21, p726-8

Price, Donald D. and Aydede, Murat (2006) "The Experimental Use of Introspection in the Scientific Study of Pain and its Integration with Third-Person Methodologies: The Experiential-Phenomenological Approach" In *Pain: New Essays on Its Nature and the Methodology of Its Study*, Murat Aydede (Ed.), Cambridge, Massachusetts: MIT Press

de Quincey, Christian (2000) "Conceiving the Inconceivable?" (discussion of N. Humphrey's target article) in *Journal of Consciousness Studies* 7, no. 4, 2000

Reid, Thomas (1764) *An Inquiry into the Human Mind.*

Riekki, Tapani; Lindeman, Marjaana; & Lipsanen, Jari (2013) "Conceptions about the mind-body problem and their relations to afterlife beliefs, paranormal beliefs, religiosity, and ontological confusions" *Adv Cogn Psychol*. 2013; 9(3): 112–120

Revonsuo, Antti (2010) *Consciousness: the science of subjectivity* Psychology Press

Robbins, Stephen E. (2013) "Form, Qualia and Time: The Hard Problem Reformed" *Mind & Matter* Vol. 11(2), pp.154-55

Rovane, Carol (2000) "Not Mind-Body but Mind-Mind", in discussion of N. Humphrey's target article in *Journal of Consciousness Studies* 7, no. 4, 2000

Rupert, Robert D. (2015) "Embodiment, Consciousness, and Neurophenomenology: Embodied Cognitive Science Puts the (First) Person in Its Place", Sec3.4, Preprint for *Journal of Consciousness Studies* 22(3-4) February 2015

Ryle, Gilbert (1949) *The Concept of Mind* Hutchinson University Library

Sayre, Kenneth (1976) *Cybernetics and Philosophy of Mind* Routledge and Kegan Paul

Sebastien, M. A. (2016) "Functions and Representations" preprint version of chapter for *Phenomenology and the Cognitive Sciences* Springer

Sellars, Wilfred (1963) "Philosophy and the Scientific Image of Man" from Sellars *Empiricism and the Philosophy of Mind* Routledge and Kegan Paul

Shapiro, Larry (2007) "The Embodied Cognition Research Programme" *Philosophy Compass* 2/2 (2007): 338–346

Shepard, Roger N. and Hut, Piet (1998) "My Experience, Your Experience, and the World We Experience: Turning the Hard Problem Upside Down" in Hameroff et al.

Singer, W. (2015). "The Ongoing Search for the Neuronal Correlates of Consciousness", sec 3. In T.Metzinger & J.M.Windt (Eds). OpenMIND:36(T)

Singh, Manish and Hoffman, Donald D. (2013) "Natural Selection and Shape Perception: Shape as an Effective Code for Fitness" An invited chapter for the Springer book Preprint of chapter for *Shape Perception in Human and Computer Vision: An Interdisciplinary Perspective* Ed. by Dickinson, Sven and Pizlo, Zygmunt. Springer

Sloman, Aaron (1994) "Towards a General theory of Representations" (preprint)

Sloman, Aaron (2004) "Information-Processing Systems in Nature"
http://www.cs.bham.ac.uk/research/projects/cogaff/sloman-information-nature.pdf

Sloman, Aaron (2007) "Why Some Machines May Need Qualia and How They Can Have Them: Including a Demanding New Turing Test for Robot Philosophers"
http://www.aaai.org/Papers/Symposia/Fall/2007/FS-07-01/FS07-01-002.pdf

Sloman, Aaron (2011) "What's information, for an organism or intelligent machine? How can a machine or organism mean?" (preprint)
http://www.cs.bham.ac.uk/research/projects/cogaff/sloman-inf-chap.html

Sloman, Aaron and Chrisley, Ron (2003) "Virtual Machines and Consciousness" preprint for *Journal of Consciousness Studies*, 10, No. 4–5

Sloman, A. and Chrisley, R.L. (2005) "More things than are dreamt of in your biology: Information-processing in biologically inspired robots" preprint for *Cognitive Systems Research* Vol 6 Issue 2, June 2005, pp145-174

Smith, Brian Cantwell (1996) "The Foundations of Computing"
http://ageofsignificance.org/people/bcsmith/print/smith-foundatns.ps

Solms, Mark (2014) "A neuropsychoanalytical approach to the hard problem of consciousness" offprint from *Journal of Integrative Neuroscience*, Vol. 13, No. 2

Spencer, H. (1890) *Principles of Psychology* 3rd Edition, vol 1, London: Williams & Norgate

Sterpetti, Fabio (2016) "Models, Brains, and Scientific Realism"
In L. Magnani & C. Casadio (eds.), *Model Based Reasoning in Science and Technology. Logical, Epistemological, and Cognitive Issues*. Springer 639-661
http://philpapers.org/archive/STEMBA-5.pdf

Stone, M.W.F. (2000) “The soul’s relation to the body: Thomas Aquinas, Siger of Brabant and the Parisian debate on monopsychism” in *History of the Mind–Body Problem* Edited by Tim Crane and Sarah Patterson Routledge

Stratton, George M. (1896) “SOME PRELIMINARY EXPERIMENTS ON VISION WITHOUT INVERSION OF THE RETINAL IMAGE” Read at the Third International Congress for Psychology, Munich, August, 1896.

Talbott, Stephen L. (2004) “Reframing the Mind-Body Problem: An Exercise in Letting Go of Dualist Assumptions” http://natureinstitute.org/txt/st/mqual/epist.htm

Thompson, Evan; Palacios, Adrian; and Varela, Francisco J. (1998) “Ways of Coloring: comparative color vision as a case study for cognitive science” in *Language form and Language Function*, edited by Frederick J. Newmeyer. MIT Press

Tomasik, Brian (2014/updated 2016) “Is There A Hard Problem of Consciousness?” [online]

Tsou, Jonathan Y. (2013) “Origins of the Qualitative Aspects of Consciousness: Evolutionary Answers to Chalmers’ Hard Problem” preprint for chapter in Liz Swan (ed.), *Origins of Mind* Springer 259-269

Turok, Neil S. (2015) “The Astonishing simplicity of Everything” public lecture, https://www.youtube.com/watch?v=f1x9lgX8GaE

Tye, Michael (2006) “Absent Qualia and the Mind-Body Problem” *Philosophical Review*, Vol. 115, No. 2

Van Gulick, Robert (2001) “Reduction, Emergence and Other Recent Options on the Mind/Body Problem” *Journal of Consciousness Studies*, 8, No. 9–10, pp1-34

Varela, Francisco; Thompson, Evan; & Rosch, Eleanor (1991/1993) *The Embodied Mind: cognitive science and human experience* MIT Press

Velichkovsky, Boris M. and Pannasch, Sebastian (2000) “In search of the ultimate evidence: The fastest visual reaction adapts to environment, not retinal locations” in commentary on O’Regan and Noë, *Journal of Consciousness Studies* 7, no. 4, p1008

Westerhoff, Jan (2016) “The Consequences of living in a virtual world” 2015 version of paper for *Erkenntnis* June 2016, Volume 81, Issue 3, pp 507-528

Wright, Wayne (2007) “Explanation and the Hard Problem” Preprint for *Philosophical Studies* (2007) 132: 310-330]