Journal of Integrative Neuroscience, Vol. 9, No. 2 (2010) 153–191
© Imperial College Press
DOI: 10.1142/S0219635210002202

Imperial College Press
www.icpress.co.uk

Commentary

A NATURAL SCIENCE APPROACH TO CONSCIOUSNESS

EDWIN R. LEWIS*
Department of Electrical Engineering and Computer Science
University of California at Berkeley, CA 94720-1770, USA
lewis@eecs.berkeley.edu

RONALD J. MACGREGOR
Department of Aerospace Engineering Sciences
University of Colorado, Boulder, USA
ronaldjmacgregor@msn.com

Received 3 August 2009
Accepted 30 April 2010

We begin with premises about natural science, its fundamental protocols and its limitations. With those in mind, we construct alternative descriptive models of consciousness, each comprising a synthesis of recent literature in cognitive science. Presuming that consciousness arose through natural selection, we eliminate the subset of alternatives that lack selectable physical phenotypes, leaving the subset with limited free will (mostly in the form of free won't). We argue that membership in this subset implies a two-way exchange of energy between the conscious mental realm and the physical realm of the brain. We propose an analogy between the mental and physical phases of energy and the phases (e.g., gas/liquid) of matter, and a possible realization in the form of a generic resonator. As candidate undergirdings of such a system, we propose astroglial-pyramidal cell and electromagnetic-field models. Finally, we consider the problem of identification of the presence of consciousness in other beings or in machines.

Keywords: Consciousness; brain; philosophy of mind; the demarcation problem; working memory; free will; free won't; astrocytes; pyramidal cells; electromagnetic fields; resonator; binding rhythm.

1. Introduction

The historical roots of western rational thought regarding consciousness and its study go back to pre-Socratics and Socratic giants of Athens. Much medieval writing addresses consciousness as one with the soul. The foundational writings of Descartes (1596–1650), and then Locke, Berkeley, and Hume can be seen as a rational secular

*Corresponding author.

renaissance in philosophy of mind which culminates in the watershed classical formulations of Kant (1724–1804). Although they were immersed in an era of emerging empiricism, these thinkers took largely intuitive approaches to their deliberations about consciousness and mind. Nonetheless, much of their thinking, especially that of Kant, rings true even to modern ears and is indeed actively extended in twentieth century thought in writers like Collingwood, Husserl, Korner, Putnam, Strawson, and Wittgenstein [1, 23, 25, 35, 37].

However, modern Philosophy of Mind has been said [25] to grow from: the writings of Wundt ([73], who founded the first experimental psychology laboratory, 1879); Brentano ([8], who in 1874 resurrected Aristotle's and Aquinas' notion of *intentionality* as the heart of the mental, which greatly influenced Freud [27]); and William James ([36], in 1890, whose monumental study of the entire range of phenomena and issues raised by the coupling of the mind with physiology firmly established the field itself throughout the western world). In the work of these pioneers we frequently see strong elements of empiricism being combined with intuition. This stream evolved through the last half of the twentieth century in the hands of major figures like Searle [69] and Chalmers [10, 11], including significant contributions from many others [13, 14, 20, 22, 30, 44, 59].

In this process and into our contemporary scene, the fields of Cognitive Science (with its top–down functional approach to Artificial Intelligence and Psychology) and Neuroscience (a strong empirical child of Medicine and Biology) emerged as the major forums of consciousness study. The potential for revealing a bottom–up undergirding of consciousness through Neuroscience is deepened by Neuroscience's relation to sister biological disciplines of Neurobiology and Molecular Biology, both of which also developed dramatically during this time.

For many years, the Center for Consciousness Studies at the University of Arizona has sponsored international conferences titled "Toward a Science of Consciousness". We take a somewhat narrower approach in this commentary, seeking paths toward a *natural science* of consciousness. We begin with a description of what we see as philosophical bases of natural science and the limitations they impose on what we can expect, ultimately, in terms of an understanding of consciousness. These limitations lead us next to a collection of descriptions (descriptive models) of consciousness. These are top–down structures that include critical elements drawn from Baddeley [3, 4], Baars [2] and others in the field of cognitive science. A key assumption — namely that consciousness has arisen through natural selection — favors a particular subset of the descriptive models, and those in turn bear implications regarding association of molecular biology and of energy with consciousness.

In Secs. 4 and 5, we discuss the hypothesis of a bottom–up generic biomolecular resonator model for consciousness and several particular putative realizations of it, including interactions between astroglial cells and pyramidal cells as well as electromagnetic fields in the brain. Finally, we return to our philosophical considerations — asking how the natural scientist would infer the presence of consciousness in an animal or a machine.

2. The Boundaries of Natural Science

Newton's *Principia* is frequently taken to mark the birth of modern natural science. Gilbert, Galileo, Harvey and others had established empiricism as the foundation of natural science. Use of mathematics to describe, concisely, the observations of the empiricist also had been demonstrated, especially by Descartes and Kepler. In the *Principia*, we see mathematics being used to endow natural science with explanatory power, through reductionism. Kepler's laws (descriptions) of planetary and lunar orbits were *explained* in terms of the laws (descriptions) of motion derived from observations (by Galileo, Descartes and Newton) on small objects in the laboratory, extended through a law (description) of gravity. What Newton did was demonstrate that the descriptions of motion and gravity are consistent with the descriptions of orbits. What he did not do, in our opinion, was demonstrate that nature obeys the laws of mathematics. We take the real numbers and the laws of mathematics to be human inventions; we would not look to those inventions in themselves for understanding of nature. They are, however, powerful tools for projecting empirical observations of nature into broad implications — as demonstrated by the works of great mathematical physicists, such as Newton and Gibbs. Newton also did not demonstrate that reductionism is a road to ontology. He defined the mass of a particle, for example, by the effects of force upon it; he defined force by its effects upon mass. Modern physicists might restate Newton's first and third laws of motion in terms of conservation of momentum, the second law as equating force with transfer or flow of momentum. If we follow the concept of *particle* into modern physics, we encounter descriptions involving superposition states, quantum entanglement and violation of Bell's inequality. Physics has not given us the ultimate nature of the particle in itself, mass in itself, force in itself or momentum in itself.

2.1. *The premises*

Premise A1. *Natural science concerns nature itself, and not the inventions of humankind, such as mathematics.*

Premise A2. *Natural science is, quintessentially, empirical.*

Premise A3. *Explanations in natural science are based on hierarchies of organization, whereby the objects and phenomena at one organizational level become the components of objects and phenomena at the next level.*

Premise A4. *Explanations in natural science are based on empirically derived descriptive models of natural objects or natural phenomena, frequently imagined in idealized circumstances.*

Premise A5. *Explanations in natural science comprise demonstrations of consistency between descriptive models at one level and those at the next level.*

Premise A6. *Natural science comprises descriptions of nature, based on human observation. It does not deal with ultimate reality or with the existence or ultimate natures of things in themselves. It is not ontology.*

Our view of natural science conforms to instrumentalism and nominalism. We believe that our descriptive model of natural science is self-consistent, a natural science of natural science itself, unencumbered by unnecessary premises.

2.2. *The demarcation problem*

In its general sense for natural science, *hypothesis* would include any empirically derived law (descriptive model) of nature such as Hooke's law (of elasticity), Newton's three laws of motion and the inverse-square law of gravity. The current criteria for such *descriptive hypotheses* to be acceptable elements of natural science include: (1) a basis in *empirical* evidence, (2) *contingency* (acceptable hypotheses must be neither true nor false by pure logic or definition alone), (3) *testability* through doable experiments or observations. All five laws met all three criteria.

The explanatory part of the *Principia*, namely demonstration that elliptical orbits obeying Kepler's laws arise from the three laws of motion combined with the inverse-square law of gravity, is another form of hypothesis — an *explanatory hypothesis*. It seems that inevitably, in order to be more than an alternative descriptive hypothesis, the explanatory hypothesis must be based on a synthesis, comprising two or more descriptive hypotheses at one organizational level (e.g., the level of the individual particle), predicting (demonstrating consistency with) observed behavior (in the form of one or more descriptive hypotheses [descriptive models]) at the next higher organizational level (e.g., at the level of two particles).

The major pitfall in this enterprise seems to be affirmation of the consequent. Among the set of all explanatory hypotheses that could be consistent with observed behavior at the next higher organizational level, which one is most applicable to the situation at hand? This is the issue addressed by Ockham's razor and the prime criterion for acceptability among explanatory hypotheses — *parsimonious basis in descriptive hypotheses that are acceptable and appropriate*. Ordinarily, in a discussion of the nature of natural science, this theme would require considerable elaboration. For the purpose of identifying a path toward a natural science of consciousness, however, it is not yet a relevant issue. All we need at this stage is that our explanatory hypotheses be based on descriptive hypotheses that are acceptable (according to the three criteria in the previous paragraph), and that they be testable. Parsimony and the appropriateness of the descriptive hypotheses (i.e., how well they apply to the perceived situation at hand) can come later.

A final caveat for this section derives from premise A6, which is central to our view of natural science. No matter how successful we are in pursuing a natural science of consciousness, we do not expect to find revelations about the ultimate nature of consciousness in itself, or about the ultimate nature of the elements that provide its foundation.

2.3. *Penrose, testability, and natural science of the conscious brain*

Penrose [59] has argued that human thought cannot be based on any current scientific principles, and that it transcends formal logic, especially that of discrete computer-like algorithmic processing which constitutes the Computational Neuroscience model of the brain. If, as Penrose conjectures, the presently-known principles of physics (which we would interpret as descriptive models reflecting the observations made so far on the basic properties of matter) are inadequate to account for consciousness, then what he requires us to do is to think beyond the currently available empirical evidence. As long as this exercise involves positing physical properties and laws that are testable by observation, it falls within our definition of natural science. If, on the other hand, it were based on premises that are inherently not testable by physical observation, it would fall outside the boundaries of natural science. It nonetheless could remain well within the boundaries of axiomatic science (mathematics and its kin). In mathematics, one is free to posit anything he wishes.

We should not be too quick to make judgments about inherent lack of testability, however. When it was made, the Einstein–Podolsky–Rosen argument for local hidden variables might have been considered inherently untestable, and thus beyond the realm of natural science — but nearly 30 years later, John Bell showed that it was testable (and it failed). On the other hand, it seems that if we accept the existence of the mental realm as compelling evidence against physicalism, i.e., against the notion that everything mental has a basis in the material, then we have taken refuge in a defined irrefutability, making natural science powerless and the *explanatory gap* (Sec. 3.5) permanent. That is not an acceptable position for the compulsively curious natural scientist. The issue of testability is addressed further in Sec. 6.

3. Top–Down Descriptive Models of Consciousness

3.1. *Foundational hypotheses*

Defining our path by premises A1 through A6, the three criteria for acceptability of descriptive hypotheses, and the two criteria for acceptability of explanatory hypotheses, we can begin our journey toward a natural science of consciousness. In the *Principia*, the destination was defined by Kepler's descriptive model of orbits. To define the destination in our case, we require a descriptive model of consciousness itself. We begin with a set of five foundational hypotheses. The first is a descriptive hypothesis based on empirical data from a broad range of species, richly and extensively supported in the literature of natural science.

Hypothesis 1. *The brain receives and processes sensory inputs, constructs, initiates and controls externally directed motor activities, and modulates some internal motor activities and some hormonally based physiological states.*

The remaining four are descriptive hypotheses based on introspection. It is prudent to recall that introspection may be filtered (confabulated and deluded) by what Michael Gazzaniga has labeled the *left-brain interpreter* [29].

Hypothesis 2. *The human brain initiates and carries out cognitive processes.*

Hypothesis 3. *The human brain is aware of some aspects of sensory processing, motor initiation and control, modulated physiological states, and cognition.*

Hypothesis 4. *Consciousness is, at a minimum, a manifestation of that awareness.*

Hypothesis 5. *The language of awareness consists of perceptions and feelings.*

3.2. *Perceptions and the first problem of consciousness*

Although the literature is rich with data concerning the neural correlates of perceptions, and the relationships between neural deficits and deficits in perceptions, presently we have no understanding of how awareness arises from neural structures. It is this connection, between the mental and the physical, that we seek; and it is this connection that presently seems beyond the reach of quintessentially empirical natural science. It is the first problem of consciousness. Through the presumption, based on individual introspection, that our fellow humans possess awareness and perceptions, we can conduct empirical studies of cognitive processes (cognitive science) and we can bridge the gap between the physical world and perceptions with empirical psychophysical studies. But the gap itself remains. And our descriptive model of consciousness is constrained by it — to be strictly in the mental realm. The elements of the model are simply the elements of perception, including that most basic element — *attention*.

Elementary perceptions include exteroceptive qualia — perceptual qualities arising from the external senses: among these are olfactory qualia (e.g., the scent of the banana), gustatory qualia (e.g., sweetness), tactile qualia (e.g., warmth), visual qualia (e.g., redness), and auditory qualia (e.g., loudness). They also include somatic or proprioceptive perceptions, perceptual qualities arising from internal sensors such as those related to balance and kinematics, bowel and bladder, chewing and swallowing, and ventilation. Among the elementary feelings are those of pain, pleasure, anger, fear, joy, sadness, lust, satisfaction, pride, guilt, envy, hunger, and satiation.

3.3. *Deluded rider model*

A key element that we have left out of the model is *volition*. Hypotheses 2 and 3 imply that the brain exercises decision-making processes as it initiates and maintains physical and mental actions. Among those actions we include (under Hypothesis 3) shifting and maintenance of attention. For now, we take the brain's awareness of decision-making processes and action initiation to be the perception of volition. Libet's work along with the corroborative work by others implies that, at least in some instances, the perception of volition occurs well after initiation of the action in question by the physical brain [44–47]. This makes tenuous the notion of free will executable by consciousness (awareness) itself. In that regard, our model so far conforms to Francis Crick's *Astonishing Hypothesis* [14]. It could, in fact, be described as a *deluded rider model*, with the delusion presumably provided by something akin to the left-brain interpreter.

What we have described so far seem largely to be here-and-now perceptions, and thus fit Damasio's category of *core* consciousness [17]. His sister category is *extended consciousness*, extended spatially beyond here and temporally to either side (past or future) of now. Another way to organize perceptions would be *innate* versus *learned* (derived through experience). Thus, in a Kantian sense, there may be an *a priori* vocabulary of perceptions, which includes many of those in our list of elementary perceptions. On the other hand, some relatively elementary perceptions, even early in the life of a human baby, must be part of an experience-based vocabulary. A fundamental element of this added vocabulary would be the perception of familiarity. Eventually, a vocabulary of associations develops — certain olfactory qualia become perceptual labels for specific food items, certain auditory qualia (e.g., pitches and timbres) become perceptual labels for the voices of specific people, and so forth. This, in turn, suggests another category of perceptions — those drawn from memory.

If we follow a common, current practice of dividing memory into *procedural memory* and *declarative memory* then we might split this last category of perception into two corresponding subcategories. Although a key property of procedural memory is its subconscious nature, awareness of the decision to carry out a procedural action sequence implies perception of any attempt to initiate that sequence. Thus, each callable action sequence, including adjustments of ongoing sequences, must be represented in the vocabulary of perceptions. Consciousness in the deluded rider model implies awareness, not volition. One might imagine the representation of an action sequence as its calling code, and consider it to be an element on an action tablet. In a free-will model, the tablet would embody the repertoire of action sequences that have been practiced and committed to procedural memory, ready to be called at will. In the deluded rider model, it would embody the vocabulary through which the conscious mind is made aware of the fact that a particular sequence, or set of sequences, has been called.

Consciousness (awareness) seems to have very limited momentary capacity. Moment by moment, we are able to attend very few elements of our vocabulary of perceptions — perhaps just one at a time. Mental structures such as short-term auditory memory and short-term visual memory allow us to maintain order in our spatial and temporal sequencing of these elements. These are the mental structures that embody Damasio's *extended consciousness*. Following the lead of Baddeley and his colleagues [3, 4], we incorporate these structures into the single mental entity — *working memory*. The vocabulary of working memory grows profoundly with experience — with each element of that added vocabulary representing increasingly complex situations that would have been represented in the inexperienced being by a number of vocabulary elements far too large to attend at once. This sort of chunking implies strong and rapid interplay between working memory and long-term memory. Its obvious adaptive value makes one wonder how widespread it must be among animals with centralized nervous systems. It also makes one wonder how the results of certain human action sequences might be represented in working memory. For example, when drawn into consciousness from memory, are simple human sounds represented by elements of auditory qualia, by elements of the action tablet, or by a combination of both? In other words, are they sub-vocalizations; and, if so, how are sub-vocalizations represented in terms of perceptions? This is an example of a set of questions about mental phenomena that might be addressable through observation of neural correlates (e.g., by functional imaging). There are many others — including immediate volition and the Libet effect [44], precedence of somatic or physiological effects and the Iowa gambling task [6].

3.4. *Cognition and working memory*

We have arrived at awareness of cognition. We can be aware of the initiation of cognitive processes and of decisions made during those processes. We can be aware, for example, of decisions to search long-term memory for a particular item (e.g., the answer to question posed explicitly or implicitly). We can be aware of the processing of items in working memory. We can be aware, for example, of our brain creating new prose or new logical relationships from old perceptual vocabulary elements. Again, following the lead of Baddeley and his colleagues [3, 4], we take working memory to be the mental substrate for all of our awareness of cognition. Again, in our deluded rider model, consciousness includes perception of volition, but there is no volition executable by consciousness itself. In the deluded rider model, consciousness is not an agent of any action, not of cognitive action, not of motor action.

Baddeley's descriptive model of working memory, however, includes, as its central elements, the *central executive*, which focuses attention, and the *episodic buffer*, which, among other things, is the seat of chunking. It seems that Baddeley associates conscious control (volition) with the central executive, and places that element in control of the episodic buffer. Thus, "retrieval from the buffer is assumed to operate principally through the process of conscious awareness" [3]. Energization and control of any mental element is beyond the capability of our deluded rider model of consciousness. Based on our initial premises, we presume that all mental elements, including awareness itself, arise from activities of physical (neural) substrates. In Baddeley's model we now confront the possibility of the inverse, activities of physical substrates being modulated by mental activities. This is a huge leap. It seems much simpler (parsimonious) to imagine the physical brain modulating its own activities — physical acting on physical, in some sort of deterministic or stochastic manner, with awareness trailing along as the deluded rider. Thus the physical (neural) activities that give rise to a particular mental construction give rise as well to what we perceive to flow from that construction. And it would flow in precisely the same way with or without awareness. Furthermore, where awareness was normally present, lack of awareness would imply lack or malfunction of the physical (neural) activities underlying that awareness. Therefore, correlations of brain or mental malfunctions of any sort with deficits in awareness do not imply that awareness itself is a causative agent. Instead, they simply affirm the consequent. It seems that it is this possibility that gave rise to Descartes's automatons and to the lack of a broadly accepted test of consciousness in animals and machines. Isn't it always possible to embed, in a mindless machine, algorithms designed to pass any test of consciousness? We shall return to this last question.

In foundational hypotheses 3 and 4 we already seem to have bypassed the arguments of the previous paragraph: *The human brain is aware of some aspects of sensory processing, motor initiation and control, modulated physiological states, and cognition. Consciousness is, at a minimum, a manifestation of that awareness.* We take the physical brain to be the substrate of awareness. Although it might be tempting to argue that awareness (consciousness) therefore must, by definition, feedback to physical states in the brain, this would remove contingency and thus take us beyond the borders of natural science. Staying on our path, we must consider the problem empirically.

3.5. *Evolution and the second question of consciousness*

It is widely held by modern biologists that consciousness arose through evolution, in response to selective pressures. Cognitive scientists have constructed thoughtful lists of the selective values that accompany consciousness. These might be considered speculative; but, according to our foundational hypotheses, *awareness* does exist. It may, as some have proposed, be a fundamental property of matter itself, with proto-awareness being an attribute of elementary particles. At present, such a descriptive hypothesis would fail to meet any of the three criteria (empirical basis, contingency, testability) for acceptability in natural science. In fact, we know of no alternative to the evolutionary hypothesis that falls within the boundaries of natural science as we have modeled it in Sec. 2. Awareness arising through evolution is an explanatory hypothesis. The fundamental elements of evolution — heritability and susceptibility to selection, on the other hand, are hypothetical attributes of organisms and thus are descriptive hypotheses. They also meet all three criteria for acceptability of descriptive hypotheses.

Accepting an evolutionary basis of consciousness as our working hypothesis, we immediately face the question — "How could evolutionary pressures have been exerted on a deluded rider?" What are the deluded rider's physical phenotypic consequences? It seems that here we would be compelled to accept the premise that awareness, in itself, in order to be selected by evolution, must be a causative agent capable of modulating physical activity. Somehow, the brain must be capable of utilizing its awareness to modify its own physical activities in a selectively advantageous way. If we subscribe to the evolutionary argument, then the second question of consciousness is "How is this accomplished?" How does the brain translate awareness back into physical (neural) action? Somehow, the physical brain gives rise to mental activities and then is able to read and utilize the results of those activities. With respect to freedom of will, the rider may still be deluded; but the evolution hypothesis seems to require consciousness to be connected to the physical brain by a two-way street. The nature of that two-way street is the 21st century's mystery of mysteries, the "explanatory gap". It will be important for us to keep in mind the translational nature of the street — along one side (P $\rightarrow$ M) it translates physical (neural) activities to mental activities; along the other side (M $\rightarrow$ P), it translates mental activities to physical activities.

If we presume that the process of focusing attention has physical underpinnings, then conscious control of attention is a member of the set of alternative hypotheses regarding the traffic on the M $\rightarrow$ P side of the street. This property of Baddeley's central executive [3] is potentially refutable via something akin to the Libet test [46]. It certainly qualifies as a valid hypothesis in a natural science of consciousness. It does not, on the other hand, appear to provide much of a selectable trait by itself. It could allow a deluded rider to explore the workings of the physical brain, temporarily manipulating the physical switches of attention — without altering the physical brain in any other way.

3.6. *More about working memory and the theater of consciousness*

It would be inappropriate here to review extensively the best current descriptive models of working memory or theater of consciousness. Such reviews are already available [2, 3]. It is appropriate to point out here, however, that these models are products of extensive empirical studies, and that they continue to stimulate such studies — in the mental realm itself and in its neural correlates. Thus they are foundational for a natural science of consciousness. For the most part, these studies have worked at both ends of the two-way street but have not addressed directly the nature of the street itself. They do, however, address the destinations of the two sides and the control of traffic.

It seems that the nature of a perceptual quale is determined by the region of the sensory brain being stimulated. In other words, there seems to be a place principle or labeled-line principle of some sort, analogous in some ways to that in the cochlea [26]. One finds evidence for this, for example, in the celebrated stimulation studies by W. Penfield and in the visual-prosthesis studies of G. S. Brindley. One finds it as well in recent studies of synesthesia (e.g., [15]) and phantom limbs [68]. When a particular region becomes dysfunctional, not only are its associated qualia lost to conscious mind, but even the knowledge of their existence (that is, the fact that they are missing) is often gone as well — *out of physical brain*, *out of mind* (e.g., [53]). This suggests that the missing pieces of the vocabulary of perception are embodied in and only in the missing pieces of the brain. It also suggests that the physical source of the traffic that is emerging at any time from the P → M side of the street is at least partially controlled from the M end of the street; and that this control is based, physically, on a labeled-line principle. This in turn suggests that the P → M side of the street has many lanes.

Baddeley's revised model of working memory comprises five elements, their linkages to one another, and their linkages to long-term memory (LTM). The uppermost of these, the central executive, was originally based on Norman and Shallice's *Supervisory Attentional System* [57]. In Baar's theater of consciousness, it presumably would be the fellow aiming the spotlight [2]. The attention of all elements is on the spotlight's target, and it is the central executive that maintains the target or shifts to a new target. The previous paragraph suggests that each target corresponds to a particular region of the physical brain. Thus one might view the central executive as a traffic cop stationed at the M-end of the street, allowing traffic to emerge from just one chosen lane at a time and attempting to divert the traffic in the other lanes — perhaps into some subconscious temporary parking area. Occasionally, a bolus of traffic from another lane will ignore the cop and emerge into consciousness, perhaps compelling the conscious mind to change the instructions to the cop. The behavior of the cop seems to change according to the circumstances. For mental tasks performed under intense attentional focus, such as those associated with deep thought or concentrated motor activity — as in certain athletics, the cop will be especially hard-nosed. For routine tasks, such as motor activities controlled by procedural memory, the cop may be more flexible, allowing traffic through on the basis of who's honking their horn most loudly, or allowing flow from two or more lanes to merge and pass.

Current descriptive models of LTM are frequently divided, as mentioned earlier, into declarative memory and procedural memory. Memories, of all sorts, must be encoded somehow in physical structures and processes. Thus they have a physical existence. It seems that, when triggered or called, neural memories can be translated directly into actions — as in modified reflexive responses, or directly into internal physiological state changes — all of this without first passing into the mental realm. They can also be translated directly into the mental realm — as internal perceptions. When they are translated into physiological states, those states in turn may also be translated into the mental realm as feelings (dread, hunger, lust, etc.). Models of declarative LTM frequently comprise two parts, *episodic* or *autobiographical memory*, which is the memory of personal experiences that can be described with words or gestures, and *semantic memory*, which is memory of general concepts and knowledge, not specifically related to personal experience, but which can also be described with words or gestures. Baddeley's descriptive model of working memory incorporates linkages to both semantic LTM and episodic LTM, with associative links between the two. In that model, semantic LTM interacts bidirectionally with two elements of short-term memory (STM) — the *visuospatial sketchpad* and the *phonological loop*. His model also includes episodic LTM, which interacts directly with another element of STM, the episodic buffer. These three elements of STM interact with each other indirectly through the central executive. They evidently also interact directly.

3.7. *Action tablet*

As mentioned previously, working memory must also have direct access to an *action tablet* containing the repertoire of motor activities that we perceive to be consciously controllable. This presumably resides in procedural LTM. Being the element that accommodates both autobiographical history and autobiographical planning (e.g., planning or day-dreaming about future actions), the episodic buffer must interact directly with the action tablet. Furthermore, it seems likely that the phonological loop, which involves sub-vocalization of words, or other sounds, as they are cycled and recycled through working memory (as in rehearsal), includes both auditory perceptual elements and action elements (representing vocal articulations). The phonological loop, then, must interact directly with the action tablet. Similar interaction seems likely between procedural LTM and the visuospatial sketchpad.

It seems that the action tablet (in procedural LTM) would comprise an array of labels for elementary actions, an array of labels for chunked patterns (of such actions) that are commonly used, and a more temporary array of labels for chunked, specialized action patterns appropriate to current or recent activities in the episodic buffer (reflecting current or recent situations surrounding the person). Thus the repertoire of immediately accessible arrays of motor chunks would always be adapting to events in the buffer — getting ready what might be needed for immediate access. We do not intend to imply here anything about the locations of the physical representations of these motor chunk labels, merely that they have been rendered immediately accessible to the episodic buffer.

3.8. *Emotions*

To accommodate the observed impact of emotions (anxiety, depression, etc.) on performance of cognitive tasks, Baddeley recently introduced his fifth element, the *Hedonic Detector*. One can picture its action in terms of Damasio's notion of valences — positive or negative feelings of various intensities [16]. Thus, it would be the hedonic detector that translates physiological states into corresponding feelings. And it would be the hedonic detector that translates conscious thoughts in the episodic buffer into feelings (mentally displayed in the episodic buffer), into modulation of attentional decisions in the central executive, and into the physiological concomitants of emotions. It seems that here we have vivid examples of both the M $\rightarrow$ P and the P $\rightarrow$ M sides of the two-way street in action. While the P $\rightarrow$ M aspect appears to be a corollary of our foundational hypotheses, the M $\rightarrow$ P aspect could be an illusion — a mental reflection of purely physical interactions in the brain.

3.9. *Agency*

This brings us back to the question of volition. Accepting the presence of the M → P side of the street as a corollary of an evolutionary basis of consciousness does not force us to accept immediate agency in consciousness. As Baars and his colleagues have described convincingly [2], consciousness could be analogous to a play on a stage such as the episodic buffer, presented to the subconscious brain (LTM) at large. At any moment, the play may be based on current events, or it may be reflections and evaluations of past events, or it may be putative plans for future events, or it may be a flight of fancy, a non-autobiographical fiction, a dream of what might be, a puzzle created, a puzzle solved. And the audience would respond to this with structural (physical) adjustments. The contents of LTM would be altered. The potential selective advantages of such a system — a system capable of evaluating rehearsing, planning and scheming are obvious. Is consciousness required for such activities? We know that, in ourselves, consciousness is associated with at least a subset of each of them. Are there subsets that are carried out subconsciously? Regardless of the answer to the last question, it seems reasonable to hypothesize that at least part of the traffic on the M → P side of the street is the action of the play on LTM. This, however, does not imply immediate thought → action causality. The flow on the M → P side of the street could be slow, with its effects on the P side being dependent on accumulation. This would be the case, for the most part, with the putatively *mind-over-matter* aspects of psychotherapy and of Buddhist exercises.

Is there evidence for more immediate causality on the M → P side of the street? It seems that a widely-held model among cognitive scientists and their physiological collaborators is a physical brain making subconscious decisions based on environmental cues, and then acting on those decisions — with the conscious mind following these events as a spectator. So far, the empirical evidence supports this view (e.g., [5, 16–18, 44–47]). On the other hand, the empirical evidence provides growing support of conscious overriding of the subconscious decisions made by the physical brain (e.g., [58]). Thus, with respect to immediate causality, the conscious mind may exert "free won't" rather than "free will".

3.10. *Alternative descriptive (top–down) models of consciousness*

Perhaps simply by definition, any description of consciousness would include a fully functional P → M side of the two-way street and thus include awareness (perceptions and feelings). It seems that this alone must have some cost to the brain — with no obvious benefit in itself. Ultimately, all costs come down to that universal currency — energy. Regarding the M → P side of the street, an idealized description might include four attributes: (1) immediate control of attention, (2a) immediate control of various behaviors and physiological processes, (2b) ability to make immediate valuative judgments and decisions, and (3) ability to produce cumulative changes in the physical components of the brain (as in learning). We might call this *the fully-volitional model*. The utterly-deluded rider model, on the other hand, would include none of the four attributes. Between these two extremes lie 14 other combinations. Using our premise of an evolutionary basis for consciousness, we reject two of the 16 combinations: the model that lacks all four attributes and the model that includes attribute (1) but lacks the rest. Neither of these models includes a selectable M → P side of the two-way street (i.e., a physical phenotypic consequent of consciousness). Based on the current empirical evidence cited in the subsection on Agency (Libet effect, and the like), we lump attributes 2a and 2b into a single attribute (2) free won't (ability to veto a decision made subconsciously).

This reduces the number of combinations to six: Model I (attributes 1, 2 and 3), Model II (attributes 2 and 3 only), Model III (attributes 1 and 3 only), Model IV (attributes 1 and 2 only), Model V (attribute 2 only), and Model VI (attribute 3 only). Thus, in Model I, the M → P side of the two-way street lies in three cognitive elements: the central executive, the action tablet, and the episodic buffer; in Model II it lies in the action tablet, and the episodic buffer; in Model III it lies in the central executive and the episodic buffer, and so forth.

In all of these variants, the subconscious brain carries out most or all of the processes that we frequently attribute to consciousness — including evaluative analysis and the making of judgments, decision-making, and action-planning. Because they are carried out in the subconscious brain, we take these processes to be automatic — based on the current morphological structure and physiological state of the brain. In other words, we take them to be carried out strictly in the realm of P. The alternative seems to be a presumption of another level of consciousness — at the subconscious level, which simply would shift the mystery of mysteries to another level. If there is more than one level of consciousness (e.g., some sort of vague levels of awareness) then we choose to lump them all into the mental realm, M. In all of the models but IV and V, on the other hand, the current morphological and physiological states of the brain (the stuff of the P realm) have been modified over time by the presence of perceptions and feelings, the stuff of the M realm. In the case of Model VI (learning through consciousness, with no conscious control of attention and no free won't) and, presumably, that of Model III (learning through consciousness plus conscious control of attention, but no free won't) as well, this is the only selectable impact of perceptions and feelings.

Invoking the "theater of consciousness" analogy [2], one might imagine the physical brain (comprising morphological and physiological states) as a member of the audience, observing and heeding a stage play (conscious activity). But this surely perverts Baar's intention. Heeding seems to require that the physical brain be fluent in the language of the play, including especially the semantic implications of the language's elements. Thus it seems to imply some sort of awareness in the physical brain — in the realm of P. But awareness is stuff of M, not P. So selectable mapping from the language of awareness (perceptions and feelings) to cumulative alteration of the elements of P (morphological and physiological states of the physical brain) seems to require active participation of the elements of M (perceptions and feelings) — as opposed to passive observation of them by the elements of P. As physicalists, we would assert that each element of M must be undergirded by elements of P, namely morphological and physiological states — or state trajectories, and that, therefore, what we really require here is merely a mapping from physical to physical. But recall that finding a mapping from P → P is not our goal here; our goal is to identify putative selectability of M. Based on our evolution premise, we would reject any model in which awareness does not produce a selectable, physical phenotype. Thus, if they lacked agency in M, Models III and VI would be rejected.

In spite of its putative agency in M, Model VI (learning through consciousness, but no free won't, no control of attention) is very much the deluded rider. It is interesting to contemplate the play in the theater of consciousness from the perspective of the brain as a whole, which is aware (it has both realms, M and P). The play might be based largely on sensory input (environmental input), i.e., based on the world around with the conscious being taking part physically, and observing mentally; or it might be the brain's own creation, e.g., possibly considering alternative scenarios for judgment-making, decision-making or action-planning. Considered as a separate entity, consciousness, if it were the deluded rider of Model VI, would observe the same play and have the sense of free choice. If there were free choice, then it would be the subconscious brain that exercised it. In that sense, the brain as a whole would be deluded by the play only in the temporal relationships between its actual choice-making and the presentation of that choice-making in the play. The "delusion" of choice as presented in the play could serve to punctuate the responsibility of the brain as a whole for choices and thus facilitate learning (morphological and physiological changes) in the physical brain. A test for this might be a comparison of the somatic responses of a subject to the Iowa gambling task [6] when he or she makes the choices and when a third party makes the choices. Under Model VI, the entire M → P side of the two-way street is associated with construction of memory traces in intermediate- and long-term memory, nothing more. This would include, putatively, intermediate-term traces involved in self-arousal of feelings, such as anger or lust, and those involved in changes in physiological state, such as relaxation and self-induced drowsiness. To be selectable in this model, then, consciousness (the brain's awareness of itself) would be taken to be a means of constructing intermediate- and long-term memories.

In Model III (which includes both control of attention and production of learning), the temporal delusion is absent in the case of the central executive. Operationally, this seems to imply that the neural circuit actually initiating a shift of attention is one and the same as the circuit generating the perception of intention to shift attention, so the intention and initiation are simultaneous (neurologically speaking). That implication should be testable — with something akin to a Libet test. What Model III adds to Model VI (production of learning only) is conscious control of the play being presented to the deluded rider (consciousness, considered separately) and to the brain as a whole. There are two alternatives to Model III: free will and free won't with respect to the central executive. In the second alternative, free won't, the subconscious brain makes putative decisions about attention, but the conscious brain has veto power over those decisions. In either case, under this model, consciousness is not only a prerequisite of long-term semantic learning, but it also guides that learning (through control of attention).

All of the remaining models (Models I, II, IV, and V) include free won't with respect to behavior, which in itself is a selectable phenotype. In Models II and V, the conscious brain has veto power (free won't) over decisions by the subconscious brain regarding physical actions and action patterns, but not over decisions regarding attention. In Models IV and V, all learning is taken to be mediated by subconscious channels. In Model I the conscious brain has its greatest impact — with conscious control (perhaps limited to veto power) over the central executive (attention and attentional manipulation of the episodic buffer), conscious veto power over decisions about actions and action patterns, and learning and feelings mediated, at least in part, by consciousness.

4. Toward a Synthetic (Explanatory) Model of Consciousness

The two-way street, which we based on evolutionary arguments, implies two-way exchange of energy. We shall define the energy available to consciousness for transfer of information along the $M \rightarrow P$ side of the street to be *dispositional energy*. Energy being a physical entity, we could consider it to be strictly confined to the physical realm (P), in which case we would imagine the action of consciousness as gating or modulating the flow of energy between various locales in P. Any process of gating, however, requires expenditure of at least some energy on the part of the gatekeeper. This seems to require us to associate at least a modicum of energy with M itself. This would be the dispositional energy. And it would be this (information-bearing) energy that flows along the lanes of the $M \rightarrow P$ side of the street. What we need, then, is a form of energy that can be set aside and then transferred quickly over the brain. One such form is the energy of resonance. Spatially separated resonators tuned close to a common frequency tend to phase lock and share energy. This is accomplished through a field, such as a vibration field in an old-fashioned clock shop, or an electric field for electrical resonances.

The presumed evolutionary basis of consciousness, it having arisen in response to selective pressures, suggests that, like other phenotypic features, it is realized somehow in the physics and chemistry of life. This suggests, further, that it is associated with a genetic element, perhaps analogous to the FOXP2 gene.

5. Consciousness and the Brain

The most significant and useful result of the assumption of an evolutionary origin of consciousness is that it necessarily links consciousness to matter. This allows us to apply at least certain basic principles of modern science to its circumstances, if not perhaps totally to it directly [43, 49, 54]. Modern biology has thoroughly established that evolutionary emergence is served by physical changes in a gene, and a corresponding change in body structure(s) (usually a biomolecule) targeted by that gene. If the change is fundamentally a process, the material change is in a physical structure (usually a biomolecule) which produces that process.

Penrose [59] has argued that human thought cannot be based on any scientific principle, especially that of discrete computer-like algorithmic processing which constitutes the Computational Neuroscience model of the brain. He believes, rather, that consciousness transcends formal logic and does not follow presently-known laws of physics. We believe, as discussed in Sec. 2, that the properties of consciousness may require thinking beyond the currently available empirical evidence [32, 49, 51, 59], which enterprise could fall inside the realm of natural science as long as it is related clearly to testable observations. Moreover, it seems that one cannot conclude much about the nature of consciousness nor its relationships with testable physical laws until one gets a better feel of the relationship between consciousness and the matter of the brain, as the behavior of matter is what physical law, as we know it, describes. In this section we introduce a pathway consistent with this thinking.

Hameroff [32] and Penrose conjecture that consciousness may be a result of quantum gravity effects in microtubules of plasmids. Recent work by Poznanski [64] and others has tended to place conscious-related plasmid effects in fine dendrites of neocortical pyramidal neurons and also argue for continuous and not discrete cognitive processing, but discount quantum effects on normal neural processing. Molecular models for consciousness, such as microtubules, make sense if consciousness originates from a genetic mutation, and we consider other molecular models just below.

In the picture we develop in this section, (non-traditional) dispositional energy and consciousness transmute from some forms of traditional energy in some elements of brain matter. We see ongoing interchange and transmutation of energies consistent with overall energy conservation analogous to that of changes of state of their underlying matter. This is a path with experimentally observable implications and some current literature to support the idea as we point out just below. For example, it may become possible to identify physical proxies for consciousness in the brain which would resolve the testability question and greatly help the question of animal consciousness. This is discussed in Sec. 7.

In this we picture consciousness and dispositional energy as emerging and rising beyond their traditional physical energy origins somewhat as vapors rise beyond their warming liquid forms, or liquids beyond their solids. This thinking provides a potential path for physicists to apply or extend fundamental physical theory of energy itself to include consciousness. Such theoretical development is ultimately the best, foundational, and perhaps only way to get to the heart of the question of consciousness and brain. One might look for a corner in existing Physics or some development thereof for a description of "state changes of energy" to subsume consciousness within Physics. The fact that liquids have lesser constraint of shape than their solids and vapors have lesser constraint of extent than their liquids suggests a possible relaxation of the operations of physical laws in dispositional energy.

Poznanski and others have rejected the idea of quantum effects on neural processing because of the extremely small time periods of quantum processes. Yet, we see the generation and maintenance of dispositional energy as dependent on particular molecular processes, and consciousness as an ongoing interaction across neural, molecular, and perhaps sub-molecular processes and structure. This would seem to override at least the temporal criticisms made of Penrose's quantum theory.

We will now develop this view formally and apply it to a specific brain network.

Hypothesis 6. *Consciousness evolved by natural section within neural or neural-related matter.*

Corollary 1. *Because consciousness has been selected by natural selection, it must, therefore, entail some advantageous modulatory alteration of neural activity. This, in turn, implies that consciousness necessarily engages in partially free energy exchanges with some neural or neural-related process(es) and/or structure(s).*

We now develop this linkage towards a full model of consciousness. Such a model should satisfy the following requirements:

- Identify the nature and specifics of consciousness's relation to neural structure and processes.
- Identify and account for the P → M and M → P translations of the two-way street, including energy exchange.
- Produce sufficient enhancement of critical brain function for evolutionary selective advantage.
- Account for consciousness of core and extended figures, and differential identities within conscious experience, and its scientifically-inferred volitional effects.
- Be testable, preferably by both consensual collective introspective and technological means.

5.1. *An energy-based resonator model of consciousness*

We satisfy these requirements by: (1) localizing the P → M, M–M, and M → P translations within genetically-evolved conscious "resonator elements" in the brain and their couplings among themselves and with other associated traditional neural structures and processes; (2) interpreting their location in brain according to our descriptive model of conscious (as presented in Sec. 3 and summarized at the end of this section) by mapping into functional brain regions and neural circuits as hypotheses for continuing comparison with neurobiological evidence and experimentation; (3) specifying hypotheses of energy transformations for P → M and M → P translations within specified candidate brain resonator elements.

Hypothesis 7. *The neural foundations and the nature, actions, and properties of consciousness are all describable in terms of a system of resonator elements by which* *consciousness participates in energy exchanges with the brain, which mediate both the generation of conscious awareness and the active modulation of volitional processes in the brain. All resonator elements produce* P → M *transformation and engage in collective partially-free* M–M *selections, but some (perceptual) may produce lesser or no* M → P *transformation.*

Hypothesis 8.

(a) *The capacity for consciousness is undergirded by the physical matter of resonator elements, which are target structures of a consciousness gene.*
(b) *Consciousness itself is a dispositional form of energy (DE) which may relate to physical forms of energy as the phase state of a gas relates in a material way to its liquid phase. Conscious awareness and its selective release are parts of DE and under the partially-free regulation of home resonator DE advised by collective systemic DE.*
(c) *The functional identity of conscious perceptions and feelings is intrinsically determined by the anatomical locations of neurons which either house resonators or are connected directly to non-neural cells which house resonators (the labeled-line hypothesis).*

5.2. *The resonator element and its properties*

The resonator element itself is a brain element which we can picture as a biomolecule, somewhat akin to an enriched mitochondria or one of its enzymes, perhaps kindred with the biomolecular catalyst of photosynthesis, or like an enriched receptor molecule such as a gating molecule or other protein involving neural, metabolic, or other physiological processes, or a photoreceptor. We will first define the resonator element in generic terms, then identify specific primary candidates. Overall, the resonator element may receive and project metabolic, traditional neural, and perhaps electromagnetic energy with traditional neural and brain elements. These intrinsic receiving and projecting connections of resonators with traditional neural structures embed consciousness and its modulatory effects in the very fabric of neural and brain integrations in composite conscious–subconscious cooperative action. In between these neural connections, the resonators carry out three essential transformative processes: (a) transformational P → M generation of consciousness, (b) participation in partially-free coalitional M ↔ M construction of directives for its effector action; (c) control and initiation of M → P effector action.

(a) P → M *transformation*
The heart of consciousness is the energy exchange between physical and mental realms. Incident neural and brain energies may trigger physiological responses in resonators, be absorbed and converted into traditional internal conformational or internal dynamic forms, and some may be transformed into a unique "dispositional" form of energy (DE) definitive of consciousness, some of which is used to generate and maintain conscious awareness itself. (Consciousness itself may be one with this

dispositional energy, but we will maintain the distinction between the two for the purposes of clarity in discussions and of flexibility in applying the model.) The generation of DE and awareness comprises the P $\rightarrow$ M side of the street between the physical and mental realms in the brain. It undergirds perception. It is the fine focal point of the first great question of the field. In this we picture consciousness and dispositional energy as emerging and rising beyond their traditional physical energy origins somewhat as vapors rise beyond their warming liquid forms, or liquids beyond their solids.

This P $\rightarrow$ M transformation is likely generated by the initiation (say, by gating) and occurrence of physiological process(es) in the element (for example, enriched enzymatic action within the Krebs ATP cycle of certain brain cells) and undergirded by what we can call the structural "generative face" of the element (the enzyme itself). Whatever the base, the transformation might be seen roughly as an analog in energy to a physical phase change, say, of liquid to gas, or normal atomic structure to plasma. The transformation and the continued maintenance of consciousness would likely require continued dissipational expenditure of energy, perhaps analogous to the burning of a flame or the heat generated in an activated electric light bulb.

(b) M $\leftrightarrow$ M *coupling*

We suppose that the partially-free selections made by consciousness are produced and maintained by an intrinsic broad, easy, fast reciprocal communication amongst all participating resonators of temporary coalitions (which latter correspond to figures of consciousness). These selections, made in individual figures and compounded across the entire configuration, produce unified consciousness, systemic consciousness, and bolster ground consciousness. They culminate in a selective pattern of distribution of directives for the subsequent release of DE across individual resonators to activate their effector actions (which comprise, for the most part, modulations of the brain controls of attention). This is the heart of partially-free conscious modulatory action.

This integrative coupling and resultant action most clearly carries the fundamental obscurity of partially-free conscious action. Perhaps we can see this best in the analogy of consciousness and dispositional conscious energy as an altered phase of physical energy, say, like the generation of a gas from its liquid form. A gas is no longer constrained in self-imposed volume; its laws of constraint are lessened, given over to different balances of forces. Similarly, consciousness retains constraints of undergirding and quantity but is free from full deterministic physical rule; it may flow freely and distribute itself within its domain as it will, something as a gas can within its larger externally-imposed volumetric constraints, as compared to liquids which must remain constrained to their net volume.

DE is commensurate with physical energy and must be entered into physical conservational balances. Similarly, we imagine that DE comprises a quantifiable reservoir of energy for the maintenance and level of consciousness, its effector actions, and associated dissipational losses, and requires ongoing replenishment accordingly.

In this model, whether consciousness (and thus the mental realm itself) is physical in the sense of physical science requires further considerations.

It would seem necessary, for example, that consciousness and DE must transcend strictly deterministic constraints on traditional physical energy to insure some partial freedom for conscious modulatory control. Integration of consciousness into physics might necessarily involve an alteration of deeper constraining principles of present material physics. For example, if one structural phase of matter may transcend some of the structural constraints of another phase (as amongst solids, liquids, gases, plasma) while not violating deeper conservation laws (matter, energy, momentum), might one phase of energy (dispositional) transcend some of the energetic constraints of another phase (lawful determinism) without violating some deeper constraining principles (overall energy balances)? That is, perhaps there is some kind of internal hyperstructure of energy that differs between physical to dispositional form as molecules of solids and liquids are closely constrained to each other, while those of gases are not.

(c) M → P *transformation*
Conscious modulation of volitional action is completed by the release (according to the coalition directive as perhaps modified by the individual resonator) of some DE in the resonator to activate a physical effect in brain elements which modulate the control of attention. The M → P process is the energy transformation, and the immediate physical effect is probably biomolecular gating of a physiological process (such as transmembrane fluxes of nutrients or ions). This physical undergirding is the "effector face" of the resonator. This process is the fine focal point of the second great question of the field. It might be likened to an elemental reversion or return meltdown of phase in the phase-change analogy.

5.3. *Localization of consciousness*

5.3.1. *By structural dynamic couplings*

In our models, consciousness is localized to a local resonator region within a parent cell. The parent (resonator-containing) neuron, or, if non-neural, the unit of parent cell and the neuron(s) with which it directly connects, is the key organizational element in the structure of consciousness (PL, the "locale" of consciousness). The dynamical coupling of active PL elements into small multiunit groupings, say of the size of single cortical modules or less [21, 71] produce what we can call "modular" or "local" circuit patterns. Such smaller circuits may correspond to fragments of figures in consciousness. Larger groupings (perhaps multiple modules, or recombined patterns) within regions (say, auditory cortex) produce neural circuits of "coalitions" of neurons which correspond to conscious "figures". Yet larger groupings, formed from multiple coalitions across multiple regions, are called "configurations", and may be labeled "fields" of consciousness. Core consciousness would seem to coincide largely with regional coalitions while extended consciousness would correspond largely with multiregional configurations, although there would be exceptions and overlap of their two size ranges.

5.3.2. *By the descriptive model of Sec.* 3

The descriptive model suggests the following localization of consciousness in the brain. It is significant that much of this seems to implicate neocortical circuitry. According to the labeled-line hypothesis, localization of function is innately determined by the physical location of individual structural elements in the brain. We can describe consciousness location according to resonator elements, partner neurons, modules [21, 71], functional regions [56], and multiregional areas. Resonator elements are the ultimate source of consciousness, we may say they are its essence. When activated, resonator elements provide awareness in itself with no necessary connection to content (localization) and in this comprise a capacity for ground consciousness. When partner neurons of given activated resonators are activated, that activation establishes a minimal elemental unit of localized conscious content corresponding to the location of the partner neuron.

Higher order conscious experience of "figures" of content corresponds largely to coordinated activity patterns in neuronal "coalitions" (circuits) containing multiple active partner neurons. The content and uniqueness of the figures correspond to the locations and uniqueness of the coalitions. In neocortex, most core figures correspond to multimodular coalitions (of pyramidal neurons) within a given region of neocortex: perceptual qualia (see Descriptive Model in Sec. 3) relate to such coalitions in auditory, visual, and somatosensory cortex. Modular and submodular coalitions correspond normally to fragments of figures, but may be enhanced by focused attention.

The figures of extended consciousness correspond mostly to higher order coalitional configurations across multiple cortical regions. These include especially working memory which is driven by coalitions of pyramidal neurons in associative neocortex through recurrent reentrant loops traversing anterior athalamic temporal lobe, and cingulate and hippocampal cortex. Its projections back through Meynert's nucleus provide the connection to formation and long-term storage of semantic memories. Its limbic circuits and projections to prefrontal cortex undergird autobiographical memory. (See Noback *et al.* for basic brain circuits [56].) Such higher-order multiregional coalitions serve progressively complex abstracted integrations such as hierarchical "chunking" and associations, complex thought and planning, inferences, daydreams, intuitions, judgments, and valuations.

Non-cognitive core and extended consciousness such as feelings, moods, emotions, senses of value, significance and the like are more closely related to elemental physiological states in hypothalamic (appetitive), reticular (arousal, intralaminar thalamus), and limbic circuitry (cingulate, prefrontal limbic cortex). In all these cases conscious content is defined and localized by active partner neurons of active resonator elements.

We expect frontal cortex to undergird overall regulatory governance and volition and to effect planning, preparation, and guidance of actions by means of pre-motor and motor cortex and basal gangliar-thalamocortical loops. Conscious perception of motor behavior is mediated by resonators and coalitions in somatosensory cortex. Conscious perception of cognitive and emotional premotor activities of the action tablet is seen as mediated by corresponding resonators and coalitions in frontal and perhaps prefrontal limbic cortex.

Conscious volitional modulation of actions is seen as the resultant of attention search cycles at the locus of attention (say, the astroglial pyramidal interface in the frontal cortex). The interplay model restricts this influence to a buffered higher-order influence on motor behavior through adjustments of long-term memory. The "free won't" veto of motor action presumes a more direct influence on motor control.

Ground consciousness is consciousness without specific content, which means active resonators connected to only inactive partner neurons or to active neurons that are not bound by the labeled-line hypothesis. This latter could be so for neurons of the so-called uncommitted cortex, and perhaps other regions of the nervous system. Indeed, it may be that the labeling of the labeled lines is a phenomenon of learning. Many regions of the neural perceptual world are fixed over time (e.g. sensory systems) so that, for example, one learns to distinguish and recognize red and green, and similarly for anatomically-fixed long-term memories; whereas, other aspects of experience represented temporarily in certain non-committed zones have only a passing temporal existence and do not acquire a permanent identifiable content, while yet retaining the quality of awareness itself.

Lashley [41, 50] described cognitive consciousness as a collection of distinct trace systems which can serve nicely as an umbrella concept for the astroglia–pyramidal model described below. Individual traces are the content elements we have described as coalitions, figures, and configurations. A given trace system houses traces that relate to a given common domain of interest; for example, a work environment or a subject of interest. At any given moment one system is dominant, all others are in abeyance. All the billions of neurons of the cerebral cortex may participate in any or all of the trace systems, but they do so in different combinations and differing temporal patterns in the different trace systems.

5.3.3. *By undergirding in specified brain elements*

Our favored complete working hypothesis is that resonator elements are found as mitochondrial catalysts of the Krebs ATP cycle in astrocytes of the astroglial–pyramidal interface of the neocortex. This model, described in detail below, seems to best fit both the localization of consciousness suggested by the descriptive model and the operational implications demanded by the close association of consciousness with attention. In any case, the discussion of this model involves a number of basic considerations and implications which can be adapted to other models.

The model can be easily altered for placement of resonators solely in pyramidal cells. Other specific alternative models, including possible roles of electromagnetic fields and receptor molecules are discussed in the following subsection below. The Llinas "binding frequency" of 40 Hz [48] corresponds to a time period of 25 msec, which can be produced easily by a number of neurodynamic processes. For specificity, we identify it here with the two-transmission loop time of the pyramidal–glia feedback on glial supply of metabolites (glucose, oxygen) to pyramidal cells. This ties it closely to both metabolism and neural signaling, and, most essentially, to the presumed source of consciousness in the model.

5.4. *Molecular models of consciousness*

If we consider that consciousness is undergirded by matter, it is instructive to consider briefly the question of what particular kinds of matter this might be and why in this and not other kinds [51]. Note that the premise that consciousness evolves in particular kinds of matter provides a bottom–up approach to this question in contrast to most approaches which are based on top–down considerations of function (learning, cognition) or physical operation (fields, quantum effects, discrete spikes, etc.). The evolutionary approach strongly suggests the molecular level as central to consciousness undergirding in that consciousness would be undergirded by a specific molecule which is constructed in part by a particular gene in the DNA molecule. Questions remain as to whether the structure itself or a physical process thereof (electrical current or field, or chemical process) is the effective housing.

Two main current molecular-based models for consciousness are (a) Magistretti and Pellerin's suggestion of glutamate receptors and association with learning in cortical pyramidal cells [52] which expresses a functional association with neurochemical transmission; and (b) Hameroff and Penrose's suggestion of protein synthesis in cytoskeletal microtubules of endoplasmic reticulum [32] which expresses a functional relation with learning and with quantum fields [59].

It may likely be that Darwinian "after-the-fact" selection among random mutations according to functional advantage bests any "before-the-fact" theory of molecular structures as prone to undergird consciousness. Yet, the ideas presented in Secs. 5.1 and 5.2, which hypothesize a transformation of some biological energy to a dispositional energy of consciousness, may lead to some bottom–up guidance. This basic idea may apply to any of the three models indicated above and to other suggestions that may be made. We develop just below (c) the molecular model of a production of dispositional energy and consciousness in Krebs cycle processes of the astroglial nutritional providers of neocortical pyramidal neurons.

Further questions remain as to whether and why the effects of any of these models might be restricted to specific brain-related regions. The glutamate model is restricted according to the glutamate transmitter system which is highly appropriate for consciousness. Poznanski localizes the cytoskeletal suggestion to (d) electrical forces in the Debye layers of microtubules in fine distal regions of electronically compact neocortical neurons, in the process strongly suggesting a continuous, rather than discrete basis of all cognition [64]. At this point we can only conjecture generally that the generation of dispositional energy is associated with or capitalized on by or within only some particular cell or molecular feature or secondary activating ingredient in or around particular brain regions.

5.4.1. *The astroglial–pyramidal cell model of consciousness*

A number of observations support the fundamental hypothesis that consciousness is a property of the astroglial–pyramidal cell interface which comprises the central neural structure of the full range of neocortical circuitry. A vast literature supports the likely role of neocortical pyramidal neurons, especially the superficial layer which receives both thalamocortical and longer-range intracortical input [24, 31, 56, 69]. LaBerge and Kasevich attribute consciousness to microelectromagnetic fields of apical dendrites of superficial pyramids [38–40]. He and Raichie [34] describe cortical fMRI signals and slow cortical waves (of EPSPS in superficial apical dendrites) which coincide with consciousness. Gray and McCormick [31] describe bursting of spikes and potential oscillations in the range of 20–70 Hz in superficial pyramids, covering the same range as the Llinas 40 Hz binding frequency [48]. Many studies establish the astroglial regulation of metabolic supply to pyramidal cells [7, 9, 19, 33, 56, 66] and its special relation to synaptic transmission [52, 55].

This interface is a two-dimensional layer like a gate across the entire input face of the neocortex, thus implicating all that pass into its main functional systems, including auditory, visual, somatosensory, uncommitted, and prefrontal limbic regions. Firing pyramidal cells initiate the recurrent reentrant circuit loops of short-term auditory memory (STAM) and the visual sketchpad of working memory, and drive the recurrent thalamocortical loops of sensory perception. The preponderance of perceptions can be attributed to these circuits, and so can volitional attentive interactions with long-term memory (LTM). LTM memories themselves are formed, stored, and recalled in and by patterns of firing pyramidal cells. Subjective feelings, emotions, valuations are implicated by pyramidal cells of the prefrontal limbic area, which integrates feeling and arousal inputs from the ascending brainstem arousal and thalamic (intralaminar and reticular nucleus) reticular systems. Autobiographical episodic areas of anterior temporal lobe feed this process and conscious-related integrations of the archicortex (cingulate gyrus and hippocampus). Thalamo-basal gangliar-cortical loops control motor behavior. (See Noback *et al.* [56], for these basic regional relations.)

This model presumes that the operations of the astroglial system undergird attention, that, simplistically, the astroglial system is attention. Resonator elements in astrocytes that are the modulators of transmission of metabolites to pyramidal cells are thus the conscious modulators of attention.

This model identifies the fundamental conscious resonator element as a biomolecule like an enriched mitochondrial enzyme which operates on or in the

ATP Krebs cycle to transform some biochemical energy into dispositional conscious energy (DE), including consciousness itself. The model thus sets attention in the very energy supply of neocortical neural processing, and sets consciousness and its partially free volitional operations in the biochemical heart of this energy supply. It would be highly difficult to find higher-ranking superordinate control of brain integrations than this. It would be difficult to imagine a deeper level of control except possibly an immediate control over the life processes themselves, which nature has wisely kept to herself.

Working Hypothesis 9. *Consciousness consists of dispositional energy generated by metabolic biomolecules in resonator regions of astrocytes, which are the undergirding of attention in the human brain.*

Attention itself operates in an ongoing recurrent cycle between parent astrocytes and their partner active pyramidal neurons until a satisfactory terminal flag system is attained and then cycled into LTM. The loop time of a single cycle defines a resonant frequency which may be usefully played upon by neural dynamics to produce the kind of temporary dynamic "binding" in the figures of consciousness as suggested by Llinas and colleagues [48]. A two-transmission loop time of about 25 msec (40 Hz) is a likely number for the astrocyte–pyramid unit.

In all this, conscious dispositional energy emerges within this otherwise traditional physical picture (perhaps figuratively beyond it, energetically in some way as a gas is materially beyond its liquid stage) and in itself produces partially-free selective systemic volitional modulatory actions regarding the figures-become-flags in its present attentional field.

5.5. *Recurrent flows and the attention cycle*

Overall, attentional energy flows from patterns of distributed pyramidal physical energy demands and of coalitional conscious firing pyramidal groups to intra-astrocyte parallel physical ATP and modulative mental DE tracks for revisionist return of patterned metabolite projections to system pyramidal cells.

In fuller summary: for energy, the astroglial system projects distributed consciously-modulated physical attentional energy (in the form of metabolites, glucose and oxygen) to the pyramidal system whose subsequent neuroelectric expenditures drive selective recurrence of that attentional projection. This projection engages parallel physical ATP (molecular metabolic energy) and mental DE (conscious dispositional energy) paths within astroglial cells. Both paths arise within glial molecular metabolism and both activate biomolecular gating of metabolite release. The intervening mental DE path is regulated by partially-free collective M–M communication amongst all currently active resonator elements which includes predominantly those of all conscious figures and fields.

Information flow is in the distribution patterns of these energy fields. The guiding flow is the present pattern of pyramidal neuroelectric activity. The distribution pattern of the guiding physical energy demand which guides the physical ATP path is predominantly the present system of active pyramidal synaptic currents. The mental energy which modulates this effect at the gating effectors is partially, freely selected by the total collective DE with primary guidance of the full range of present conscious volitions, figures, and fields (identified by the present distribution of firing pyramidal cells).

Attention operates systemically by a repeated recurrent flow of energy through the collective pyramidal–astroglial circuitry. This is driven by two sources in the neuroelectric activities of the pyramidal cells. The synaptic current loops are the predominant energy expenditure of the system. These are extremely numerous, correspond to the totality of present input and recurrent synaptic transmissions and reflect accumulative graded neuronal membrane potentials. They reflect associational connections of current firing patterns as well as firing patterns themselves [41]. This produces a strong systemic "energy-demand" which must be supplied by the return flow of metabolites (glucose, oxygen) into pyramidal cells which is regulated by the astroglial cells. The present pattern of firing pyramidal cells, on the other hand, is a smaller energy demand because it reflects active cells rather than synapses (a much smaller number). Yet, firing patterns are not only the embodiments of central composite conscious structures, but also a main source of the return path for the conscious modulation of attention. We will label these composite conscious firing patterns of pyramidal cells as "flags" [49]. In this, attentional action can be characterized as a repeated recurrent modification of flag systems.

It has seemed to many, beginning with Lashley [41], that conscious figures would correspond nicely with collective configurational firing patterns. Yet, Shulman and colleagues [70] have shown that the state of consciousness is associated primarily with the greater energy demands of the synaptic current loops in superficial apical dendrites. Others have also associated consciousness with continuous volume currents or fields, some including specifically micro EM fields of pyramidal apical dendrites [34, 38–40, 45–47, 62–66]. In our astroglial–pyramidal model the resonator region of active parent astrocytes are the supporting undergirding of consciousness and all its transformations and effects, and all functional identity is intrinsic to active pyramidal cells directly supplied by these. This model would support the broader picture of the occurrence of consciousness in all active nutrient-supplied energy-demand sites. This would apply to both somatic spike-triggering regions and especially to multitudinous active dendritic synaptic sites which are richly supplied by astroglia [56, 66]. This total activity centrally includes sustained coalitional firing patterns of conscious figures (which are conscious via both firing and their suprathreshold dendritic activation), but also includes some broader, perhaps more intensely activated, associational connections and accumulations of associations. Lashley characterized this graded dendritic activity as mediating "the direction of attention". These associational synaptic activities may well play a significant role in the informational processing of the repetitive recurrent attention cycle.

In our models, a single recurrent cycle would correspond to the resonant frequency of the resonator and hypothesized "binding frequency" of consciousness, 40 Hz (Llinas and Pare [48]) and 20–70 Hz (Gray and McCormack [31]), say about 25 msec. A typical integrating time in the interplay studies is about 250 msec, an episode of ten cycles.

5.5.1. *Thinking*

Conscious modulation of volitional attentional control is perhaps most complete in the governance of cognitive searching (say, thinking) in the LTM banks of neocortical pyramidal cells. Activity patterns in these cells are selectively reinforced by increased metabolic supply from networks of astroglial cells by enhancement of successful input synapses to pyramidal cells (synapses whose firing immediately precedes firing of its parent cell). This is a Hebbian learning where post-synaptic cells learn to respond to smaller portions of successful input patterns [42, 43, 49, 50]. This could be an effective mechanism of chunking and would be a process of the episodic buffer in terms of the working memory model. Conscious modulation of this recurrent astroglial control system could be an effective means of modulatory conscious direction of thought. Variations could be used to search out optimal associative connections among the comparisons of current and recent families of active synapses as revealed within the episodic buffer.

5.6. *Electromagnetic fields and consciousness*

The participation of neuroelectrically-based electrical or electromagnetic fields (EM) seems an attractive candidate for consciousness [38–40, 45, 46, 61]. Like the generic resonator element they emerge from and project influence on traditional neural processing, they can be generated by and acted upon by the confluent normal local, regional, and global volume currents of normal neural signaling. These effects could be systemically built into brain function by the exceedingly regular network architecture and neuronal structures of many of its main circuit systems. The overall confluent EM field of the brain comprises a highly detailed picture of total brain activity as an alternative language description of composite neuronal signaling and volume currents. It also operates with practically instant unified communication. These attributes could serve as basic properties of consciousness [10, 11, 45–48, 64]. There are two problems here which would require resolution. The first is that EM is a well-known phenomenon of deterministic physics, and would be hard-pressed to produce in itself partially-free volition. The second is that EM fields are pervasive in the inert and biological realms and one doubts these are conscious. Fundamentally the EM hypothesis does not stand in itself but EM may participate significantly within a separately grounded resonator system. Indeed, any theory of attributing consciousness to a physical EM field necessarily involves a resonator model to establish the P–M association.

EM fields might participate significantly with a resonator consciousness system grounded elsewhere. For example, M $\leftrightarrow$ M communication amongst resonator elements might decompose into mirror M $\rightarrow$ P transformation driving of EM to produce (EM) P action on distant traditional brain processes. The inverse possibility is that EM could selectively trigger resonate P $\rightarrow$ M transformation generation of consciousness by enriched photoreceptor-like biomolecules in resonators specially tuned to their frequency. In this, conscious DE would be necessary for the production of conscious awareness, the partially-free pattern selective capacity not available to a deterministic physical EM field, and the gating of the modulatory alteration of the EM field at the M $\rightarrow$ P interface in the resonator. Any way to attribute any of these capacities to the EM field can be seen as a resonator model.

5.6.1. *The simplest EM model*

The simplest EM model would be a resonator whose M region consisted only of the generation and storage of DE, its continued maintenance of awareness, and its collective selective graded retransformation of itself into projected EM. In this model, consciousness would exist not directly in EM fields, yet the resonant frequency bands would always be related to consciousness because these bands would trigger and continually support its generation in resonators and the resonators would produce it. Consciousness itself would always be a property of the resonator elements. This model could be characterized as the EM field's giving or taking energy from the resonator. This could conceivably leave an indelible imprint on the departing field, similar in principle to the characteristic markings of light that reveal the chemical makeup of stars. Something like this could become an observable proxy for consciousness.

This EM field might comprise, at the P-ends of the street, an energy field that is associated with consciousness and is broadly available over the neural structures of the brain. The 40-Hz rhythm discussed by Gray and Llinas [31, 48] seems to have these properties. The energy in such a field could be augmented by local neural circuits, which could draw energy from such a field, as in the physical oscillator model described in Sec. 4. Each of these actions could be accomplished by local resonances, active or passive, tuned to the frequency of the field. One might associate some of those circuits with the neural structures underlying consciousness, and some of them with neural circuits associated with the subconscious brain viewing the play on the stage of consciousness. The composite electrical field is a highly complex spatiotemporal pattern whose intricacies could conceivably convey the composite localization information of modulated attention at the localized receptive effectors.

5.6.2. *EM coupling in the astroglial–pyramidal model*

In the basic astroglial–pyramidal model, energy flow may be largely from generator to effector sides of single resonators, whose functional meanings are defined by the location of their parent or partner neuron. DE is the medium, location is the message. An obscurity of energy flow and/or information flow may reside in the apparent unification of consciousness which in our model reduces to the partially-free collective selection (by M $\leftrightarrow$ M sharing across all active resonator elements) of the composite distribution of effector DE release across all individual resonators, as further adjusted within by each local resonator. In all this, DE itself is the capacity (energy), the experience (awareness), and the partial controller of collective DE release to gate effectors throughout the current field of effector action. In the astroglia–pyramidal model the total resonator influence on pyramidal neurons and their apical dendrites would be coupled with modulation of traditional metabolic and neural actions as well as these possible EM effects.

5.7. *Alternative resonator elements and add-in effects*

In the astroglial–pyramidal model, the M $\rightarrow$ P astroglial effector action enhances gating of glutamate receptor action whether produced directly or via metabolites, thereby suggesting a strong coupling of glutamate transmission with consciousness [52, 66].

The astroglial–pyramidal model seems sufficient in itself to cover all conscious experience and properties. Yet, other conscious elements and systems may serve as its instrument, companion, or alternative. Endymal glial cells, choroid plexi, and ventricles could support more direct involvement of hippocampal, thalamic intralaminar, and brainstem limbic influences [56].

Alternatively, mitochondrial catalysts in pyramidal cells themselves and perhaps some other neurons may serve to produce DE from an ATP process. The picture produced by direct housing of resonators in pyramidal cells is very similar in appearances to the astroglial–pyramidal model.

6. Testability (Verifiability)

We closed Sec. 2.2. with the point that affirmation of the consequent is not yet an issue in the natural science of consciousness. Were it an issue, then a follower of Popper might wish to replace *verifiability* with *refutability*. At this point, ***any*** testable, theoretical realization of even a one-way street to or from consciousness would be an earth-shaking advance, regardless of its relevance to the human brain, any brain, any nervous system. The problem is testability.

6.1. *Testability of awareness*

The *cogito* in Descartes' "*Cogito ergo sum*" involves both a computational process and awareness of that process. The computational process alone does not imply awareness. Thus, it does not imply consciousness. It is awareness that corresponds to consciousness. Paraphrasing David Chalmers [10, 11], we have no awareness meter.

Such a meter would employ the two-way street (P → M, M → P) — the very thing we seek, the mystery of mysteries. Lacking that, it seems that we can probe the putatively mental side of beings only by using the putative two-way streets of those beings themselves. What is it that we could detect on the physical side that would imply the existence of a mental side and its concomitant awareness? By what means can we distinguish a Cartesian automaton from a conscious being? It seems that what we seek is logically demonstrable *necessity* — in the sense that such and such phenomenon or phenomena on the physical side would be impossible without the presence of a mental side. This, in turn, seems to require that our descriptive model (descriptive hypothesis) of consciousness have a non-contingent attribute (logically necessary association with the physical phenomenon or phenomena in question). That places the model (and thus the consciousness meter based on it) beyond the boundaries of natural science. With descriptive models acceptable to natural science, it seems that the best we can do is find consensus in the community of natural scientists regarding pairings of experiments and results from which consciousness can be strongly inferred. This puts us at the boundary between the empirical world of natural science and a world of intuition, presumption and faith.

Because each of us is aware of his/her own consciousness, the community accepts various forms of communication as evidence of consciousness in other human beings. This makes possible the construction of *descriptive models* of human consciousness that can be validated by the community at large. It is the basis of the "natural science of consciousness" described in Sec. 3. The criteria for consciousness become increasingly controversial in clinical situations where impaired consciousness is involved (e.g., anesthesia and coma).

For non-human subjects, we believe, the issue is utterly unresolved. The presumption that consciousness is uniquely human seems to be theological, and therefore beyond the boundary of natural science. One might argue, further, that the selective advantages of consciousness are compelling — but that would be difficult to do if we cannot distinguish, by consensus-based means, between automaton and conscious being. In other words, we would, concomitantly, lack a consensus on what the automaton must lack that the conscious being might have. On the other hand, the community of biologists has established a code of ethics based on the presumption that the perception of pain is common to animals with centralized nervous systems. *Pain* has no meaning outside of consciousness, so the general presumption is that of *consciousness* in such animals. It would be, at the very least, comparable to Damasio's *core consciousness* [16–18]. The validity of the presumption, however, awaits a consensus-based test. The physiologist might use a reflex, such as a tail-pinch response, to test level of anesthesia. Sherrington, on the other hand, might have argued that the presence of reflexes is not necessarily a measure of the presence of consciousness. When the response clearly is brain-mediated, (e.g., not merely a spinal reflex in a vertebrate), the prudent physiologist certainly would add a supplemental dose of anesthetic. An acquired avoidance response such as cringing or attempting to escape in the face of an impending stimulus likely to bring pain would be pretty compelling.

There is one test — the mirror test [28] that seems to be acceptable to a subset of cognitive scientists. It is applied to human toddlers, beings already presumed by the community to have elements of consciousness or at least protoconsciousness. The mirror test probes for the presence of self-recognition — something most would say is well beyond core consciousness (e.g., ability to sense pain). In the developing toddler, self-recognition seems to arise in parallel with *autobiographical memory*, the presence or absence of which in non-humans remains controversial. Some primates have passed the mirror test [28], as have pigeons [72], magpies [67], elephants [60], and bottlenose dolphins.

The June 2008 issue of *IEEE Spectrum* was dedicated to the "Rapture of the Geeks", a presumption that consciousness will emerge at some point (the "singularity") in man-made machines — as they become increasingly complex, self-taught, and interactive with the world around them. In other words, a general faith that the conscious being will emerge from the Cartesian automaton as the latter becomes increasingly complex. How would we know? We could apply a modified Turing test (presence of consciousness, however, does not imply great intelligence). But the consensus-based acceptance of communication with another human being is based on human introspection. The community accepts tests of awareness that are based on communicative evidence when the subject is human; but it is unlikely to accept such tests for machines. On the other hand, if a robot not explicitly programmed for the situation, on being confronted by a mirror made some exploratory gestures then announced "that is me, R2D2", the community surely would be impressed. Furthermore, the experiment would, in principle, be repeatable, with repetition including constructing, programming, training, and testing the machine — not simply testing it. We say "in principle" because that sort of self-recognition may be nearly universal among members of the human species past a certain age, but it is not universal among members of any non-human primate species. In other words, a geek could faithfully synthesize a non-human primate, yet have his creation fail the mirror test. Self-recognition may be a valid criterion for something very special, but surely it cannot be the criterion for the ability to sense pain. And (core) consciousness is the *sine qua non* for that ability. Again, outside of consciousness, *pain* has no meaning. The mirror test seems to set the bar far too high, as does the Turing test.

From a physicalist's point of view, consciousness must have its underpinnings in material phenomena, presumably neural. Such a person would not expect a conscious perception (or decision) to precede its material underpinnings in time. In fact, he or she would probably expect the underlying neural activity to precede, at least slightly, the conscious consequent. This sets constraints on the interpretability of tests, such as the Libet test, that are based on relative timing of neural and conscious events. Earthshaking, of course, would be discovery and verification of a conscious event clearly preceding its neural underpinnings. Given the history of physics in the 20th century, even a physicalist must admit the possibility of such a thing — but it certainly is beyond physics as we know it — and, except for the timing experiments themselves, beyond the boundaries of natural science. Furthermore, given the complexity of the human brain, and the current limitations on detailed observation of its activities, widely acceptable verification of such a discovery would be more than exceedingly difficult, probably impossible.

It seems that in humans, consciousness is the gateway to long-term declarative memory, including semantic memory and autobiographical memory. It is the gateway for external sensory input, translated largely into semantic information, and the gateway for re-entry of the brain's own reflections and ruminations. Furthermore, it seems unlikely that, over the course of evolution, long-term semantic memory and/or its gateway suddenly arose in the human species. Therefore, demonstration of long-term semantic memory in an animal species might provide strong inference of the presence of consciousness in that species. Again, this would be well beyond the core consciousness of pain. On the other hand, acquired, brain-mediated avoidance responses to complex, perhaps subtle, contexts likely to lead to pain might reasonably be taken as declarations of semantic memories in animals.

6.2. *Testability of consciousness and the brain*

This work argues that the ultimate testability for consciousness resides in its connection with the brain and that two general pathways are foundational. The first is the presumably necessary undergirding of consciousness in the genetic structures of the brain, including likely both a gene field in the DNA and its target structure(s), probably a biomolecule. Once a consensus test of consciousness is available, one could launch the highly developed technology in each of these tracks to lead ultimately to observable correlates (proxies) of consciousness. The nature and pursuit of those studies is in the hands and minds of specialists in these fields.

The second path is the seemingly necessary association of consciousness with physical processes in the brain, which we infer must be an energy exchange of some sort. There are at least four levels where the energy connection can be pursued to the ultimate revelation of observable correlates with consciousness, each requiring the focused intelligence and skills of specialists.

The first is the demonstration of overall quantifiable impact on brain metabolism by mental behavioral tasks requiring intense usage of consciousness. Skillful cognitive psychologists can develop these.

A second level is that of combined observations of fMRI, brain neuroelectric and metabolic signals, and subjective reports of consciousness, like those of Shulman and colleagues [70] and He and Raichie [34], also available in many laboratories worldwide. The ingredient would be, again, the determination of a repeatable observable quantifiable energy correlate of consciousness.

The third level is the pursuit of the energy correlate in neural circuit patterns of likely brain regions, like pyramidal cell networks and their astroglia. Penetrating studies like those of LaBerge and Kasevich [38–40], or Newman [55] and many others can be enriched to seek quantifiable energy associations.

Fourth is the level of biomolecules and biochemistry, which, according to the evolutionary argument, is the likely home ground of consciousness. Again, advanced technology exists and the intelligence and skill of specialists are required [19].

The neural integrations of the second and third levels could be assisted by theoretical and modeling skills (See, for example, [62–66]).

7. Concluding Remarks

Throughout this essay, we have distinguished the mental (M) from the physical (P). Generally, one might presume that by *mental* we would include the memories, intellectual tools, and the like, that compose the human mind. This is not the case. What we include in M is limited to conscious perceptions and reflections, feelings and emotions. It seems clear that the human mind is a vast structure, largely subconscious, and that consciousness provides an exceedingly narrow window on that structure. It seems equally clear that the power of consciousness, and therefore of the human mind is leveraged hugely by the process of chunking — a rapid interplay between the conscious and subconscious mind. It also seems clear that the leveraging power of chunking in the human mind was further amplified, hugely, by the evolution of language. Furthermore, and perhaps most important, the capacity of working memory is expanded powerfully by the employment by humans of graphical tools, by pictures, words and symbols drawn or written — in the dust, on a sketch pad, on a blackboard. These allow human contemplations to expand with perfect retention over ranges of space and time that are vastly broader than those available in neural short-term memory alone, and they allow them to be shared in collaborations of two or more working memories.

Indeed, the highest reaches of the mind, and thus, all things human, are in the hierarchical structurings produced by chunking, with its powers to represent, place, contextualize, imagine, interplay, and thereby, understand, manage, evaluate, choose, and optimize. The leveraging power of chunking is an overriding distinction between man and other primates. The window of consciousness may be small, but the power of consciousness is also leveraged by virtue of chunking (in mind and brain) to high-level modulations sovereign to many subordinate elements, including many unconscious influences. The tradeoff between unconscious and conscious influences is at the heart of the human condition.

7.1. *Consciousness in animals*

Original prototypical prehuman consciousness might be seen as first emerging in resonator biomolecules which generate some degree of perceptions or feeling, and perhaps associated with a particular early version of a fledgling not-previously conscious "conscious-like system", say a prototype of the physical oscillator system indicated in Sec. 4. In an ineffectual deluded rider, consciousness in itself would not provide selectable advantage. Thus, either some basis of advantage would need to ensue from the energy given to awareness generation/maintenance (which rather would seem to produce a deleterious effect) or a modicum of M $\rightarrow$ P reversion from the start, or subsequent further progressive co-evolution of consciousness and neural structures need to ensue rather quickly. Subsequent co-evolution would be expected to progress quickly to eventually include extended perceptions, emotions, intuitions, understanding, arousal, flag systems, attention, unification, and so on. All such progressions would be expected in our model as increased distribution of resonator placements, couplings with both neural and resonator elements, and M $\rightarrow$ P effector connections.

It would seem that chunking may have co-evolved with further development of preexisting consciousness. The descriptive model might suggest a primary emergence associated with core consciousness (sense of pain, sharp perceptions of surroundings, food), subsequent co-evolution of short-term and working memory (extended consciousness over time and space), leading ultimately to co-evolution of long-term memory, and, especially in humans, chunking. We could characterize Damasio's thinking [16–18], for example, as a five-layered evolution along the lines of: (a) consciousness of proto-self in simple animals (wakefulness, images, attention, significance); (b) add external object relationships in fish, reptiles, primitive mammals; (c) core consciousness in higher mammals (core self, short- and long-term memory, strong here-now sense, no complex language); (d) extended consciousness in Neanderthals, chimpanzees, dolphins (limited language, autobiographical self and memory, limited sense of past and future); (e) higher extended consciousness in modern humans (adds complex language, stronger memory, conscience, artistic and scientific creativity).

The strong elevation of long-term memory, especially long-term episodic and declarative memory, in modern humans so indicative, and, we would say, of chunking, might well be taken to be dependent on consciousness and have co-evolved with it [51]. In this the presence of long-term memory would be enough to strongly imply consciousness. The co-evolutions of manifestations of consciousness in higher vertebrates from a primal genetic brain element (as, say, a resonator model) may be seen as a two-factor energetically-guided process as discussed in Ref. 51.

7.2. *Main contribution and experimental observation*

The main contribution of this work is a synthesis of major recent models of top–down functional consciousness study (cognitive science) and its melding with major recent experimental brain studies, and an hypothesized undergirding of consciousness in energy-based biomolecular resonator elements with specific example models, all within a carefully constructed overall natural science approach based on observation and carefully selected hypotheses directed towards subsequent experimental validation. The functional and brain undergirding approaches are both necessary halves of a whole which require the other for completion.

Our theoretical work is also a half which requires fulfillment by experimental explorations of its hypotheses. The future prospect of this and other brain theories is in the hands of our experimentally-gifted counterparts.

References

[1] Allison H, Bird G, Kantianism, in Honderich T (ed.), *The Oxford Companion to Philosophy*, Oxford Univ Presss, New York, 1995.
[2] Baars BJ, *In the Theater of Consciousness*, University Press, New York, 1997.
[3] Baddeley A, *Working Memory, Thought, and Action*, Oxford University Press, New York, 2007.
[4] Baddeley A, Working memory and conscious awareness, in Collins AF, Gathercole SE, Conway MA, Morris PE (eds.), *Theories of Memory*, Erlbaum Associates, Mahwah, NJ, 1993.
[5] Bargh JJ, Chartrand TL, The unbearable automaticity of being, *Amer Psychol* **54**:462–479, 1999.
[6] Bechara A, Damasio AR, Damasio H, Anderson SW, Insensitivity to future consequences following damage to human prefrontal cortex, *Cognition* **50**:7–15, 1994.
[7] Bezzi P, Gundersen V, Galbete JL, Seifert G, Steinhauser C, Pilati E, Volterra A, Astrocytes contain a vesicular compartment that is competent for regulative exocytosis of glutamate, *Nature Neurosci* **7**:613–620, 2004.
[8] Brentano F, *Psychology from an Empirical Standpoint*, 1874.
[9] Bullock, TH, Bennett MVL, Johnson D, Josephson R, Marder E, Douglas R, Fields RD, The neuron doctrine redux, *Science* **310**:791–793, 2005.
[10] Chalmers DJ, *The Conscious Mind: In Search of a Fundamental Theory*, Oxford University Press, New York, 1996.
[11] Chalmers DJ (ed.), *Consciousness and its Place in Nature*, Oxford University Press, New York, 1992.
[12] Chalmers DJ, *Philosophy of Mind: Classical and Contemporary Readings*, Oxford University Press, Oxford, 2002.
[13] Churchland PS, *Brain-Wise: Studies in Neurophilosophy*, MIT Press, Cambridge, MA, 2002.
[14] Crick F, *The Astonishing Hypothesis: The Scientific Search for the Soul*, Touchstone, Simon and Schuster, New York, 1995.
[15] Cytowic RE, *Synesthesia: A Union of the Senses*, 2nd ed., MIT Press, Cambridge, MA, 2002.
[16] Damasio A, *Descartes' Error*, Puntnam, New York, 1994.
[17] Damasio A, *The Feeling of What Happens*, Harcourt Brace, New York, 1999.
[18] Damasio A, *Looking for Spinoza*, Harcourt, Orlando, FL, 2003.
[19] Davia CJ, Life, catalysis and excitable media: A dynamic systems approach to metabolism and cognition. In Tuszynski J (ed.), *The Emerging Physics of Consciousness*, Springer-Verlag, Heidelberg, Germany, 2006.

[20] Dennett D, *Consciousness Explained*, Back Bay Books, Little Brown and Co., Boston, MA, 1991.

[21] Eccles JC, The modular operation of the cerebral neocortex considered as the material basis of mental events, *Neurosci* **6**:1839–1856. RC95. 1–4, 1981.

[22] Edelman G, *Wider than the Sky: Phenomenal Gift of Consciousness*, Basic Books, 2004.

[23] Edwards P (ed.), *Encyclopedia of Philosophy*, Macmillan Publications, NY, 1967.

[24] Elston GN, Pyramidal cells of the frontal lobe: All the more spinous to think with, *J Neuorosci* **20**:RC95, 1–4, 2000.

[25] Flanagan O, History of the philosophy of mind, in Honderich T (ed.), *The Oxford Companion to Philosophy*, Oxford University Press, NY, 1995.

[26] Fodor JA, *Modularity of Mind: An Essay on Faculty Psychology*, MIT Press, Cambridge, MA, 1983.

[27] Freud S, *New Introductory Lectures on Psychoanalysis*, 1923; *An Outline of Psychoanalysis*, 1940; see Strachey J (translated), WW Norton and Co., New York, 1949; See also *Basic Writings of Sigmund Freud*, Random House, 1977.

[28] Gallup G, Chimpanzees: Self recognition, *Science* **1267**:86–87, 1970.

[29] Gazzaniga M, *Human: The Science Behind What Makes Us Unique*, Harper-Collins, New York, 2008.

[30] Globus GG, Towards a noncomputational cognitive neuroscience, *J Cogn Neurosci* **4**:299–310, 1992.

[31] Gray CM, McCormick DA, Chattering cells: Superficial pyramidal neurons contributing to synchronous oscillations in the visual cortex, *Science* **274**:109–115, 1996.

[32] Hameroff SR, Quantum coherence in microtubules: A neural basis for emergent consciousness? *J Consc Studies*: **1** 98–118, 1994.

[33] Haydon PG, Glia: Listening and talking to the synapse, *Nature Rev Neurosci* **2**:185–193, 2001.

[34] He BJ, Raichie ME, The fMRI signal, slow cortical potential, and consciousness, *Trends in Cognitive Science* **13**:302–309, 2009.

[35] Honderich T (ed.), *The Oxford Companion to Philosophy*, Oxford University Press, New York, 1995.

[36] James W, *Principles of Psychology*, McGraw-Hill, New York, 1980.

[37] Kim, J, Problems of mind, The mind-body problem, in Honderich T (ed.), *The Oxford Companion to Philosophy*, Oxford University Press, NY, 1995.

[38] LaBerge D, Kasevich R, The apical dendrite theory of consciousness, *Neural Networks* **20**:1004–1020, 2007.

[39] LaBerge D, Apical dendrite activity in cognition and consciousness, *Conscious and Cognition* **15**:235–257, 2006.

[40] LaBerge D, Kasevich R, Microelectromagnetic fields in apical dendrites, *J Neuroscience* **26**:4046–4053, 1997.

[41] Lashley KS, *The Neuropsychology of Lashley*, Beach FA *et al.* (eds.), McGraw-Hill, New York, 1960.

[42] Lewis, ER, Neural networks, *Online Encyclopedia of Neuroscience*, 2003.

[43] Lewis ER, MacGregor RJ, On indeterminism, chaos, and small number particle systems in the brain, *J Integr Neurosci* **5**:223–248, 2006.

[44] Libet BJ, Unconscious cerebral initiative and the role of conscious will in voluntary action, *Behav Brain Sci* **8**:529–566, 1985.

[45] Libet BJ, Conscious mind as a field, *J Theor Biol* **178**:223–243, 1996.

[46] Libet BJ, *Mind Time: The Temporal Factor in Consciousness, Perspectives in Cognitive Neuroscience*, Harvard University Press, Cambridge, MA, 2004.

[47] Libet BJ, Reflections on the interaction of the mind and brain, *Prog Neurobiol* **78**:322–326, 2006.

[48] Llinas RR, Pare D, Of dreaming and wakefulness, *Neuroscience* **44**:521–535, 1991.

[49] MacGregor RJ, *On the Contexts of Things Human*, World Scientific Publishing, Singapore, 2006.

[50] MacGregor RJ, Lewis ER, *Neural Modeling*, Plenum Press, New York, 1977.

[51] MacGregor RJ, Vimal RLP, Consciousness and the structure of matter, *J Integr Neurosci* **7**:75–116, 2008.

[52] Magistretti PJ, Pellerin L, Cellular mechanisms of brain energy metabolism and their relevance to functional brain imaging, *Phil Trans R Soc (Lond): B* **354**:1155–1163, 1999.

[53] Mazur L, The self and its awareness: Genesis of psychoses, *J Integr Neurosci* **7**:551–570, 2008.

[54] McCollum G, Book Review of *On the Contexts of Things Human* by RJ MacGregor, WSPC, *J Integr Neurosci* **7**:571–581, 2008.

[55] Newman EA, New roles for astrocytes: Regulation of synaptic transmission, *Trends in Neurosci* **26**:536–542, 2003.

[56] Noback CR *et al.*, *The Human Nervous System*, Springer, 2007.

[57] Norman DA, Shallice T, Attention to action: Willed and automatic control of behaviour. In *Consciousness and Self Regulation, Advances in Research and Theory*, Vol. 4, Davidson RJ, Schwarts GE, Shapiro D (eds.), Plenum, New York, pp. 1–18, 1986.

[58] Obhi SS, Haggard P, Free will and free won't, *American Scientist* **92**:358–365, 2004.

[59] Penrose R, *Shadows of the Mind: A Search for the Missing Science of Consciousness*, Oxford University Press, Oxford, UK, 1994.

[60] Plotnik JM, de Waal FBM, Reiss D, Self-recognition in Asian elephant, *Proc Natl Acad Sci* **103**:17053–17057, 2006.

[61] Popper KR, A discussion on the mind-body problem, *Theoret Med* **14**:167–180, 1993.

[62] Poznanski RR, Introduction to integrative neuroscience, in Poznanski RR (ed.), *Biophysical Neural Networks: Foundations of Integrative Neuroscience*, MA Liebert, Larchmont, NY, 2001.

[63] Poznanski RR, Towards an integrative theory of cognition, *J Integr Neurosci* **1**:145–156, 2002.

[64] Poznanski RR, Model-based neuroimaging for cognitive computing. *J Integr Neurosci.* **8**: 345–369, 2009.

[65] Poznanski RR, Book Review of *Brain-Mind Machinery: Brain-Inspired Computing and Mind-Opening*, by Gee-Wah Ng, *J Integr Neurosci* **8**:405–409, 2009.

[66] Poznanski RR, Riera J, fMRI models of dendritic and astrocytic networks, *J Integr Neurosci* **5**:273–326, 2006.

[67] Prior H, Schwarz A, Güntürkün O, Mirror-induced behavior in the magpie (*Pica pica*) Evidence of self-recognition, *PLoS Biol* **6**:e202, 2008.

[68] Ramachandran VS, Blakeslee S, *Phantoms in the Brain*, Harper, NY, 1998.

[69] Searle JR, Consciousness, *Ann Rev Neurosci* **2**:557–578, 2000.

[70] Shulman RG, Hyder F, Rothman DL, Baseline brain energy supports the state of consciousness, *Proc Natl Acad Sci* **10**:1073, 2009.

[71] Szentagothai J, The 'module concept' in cerebral cortex architecture, *Brain Res* **95**:475–496, 1975.

[72] Toda K, Watanabe S, Discrimination of moving video images of self by pigeons (*Columba livia*), *Animal Cognition* **11**:699–705, 2008.

[73] Wundt W (1873–1874), *Principles of Physiological Psychology.*