Global Workspace Theory and Synthetic Phenomenology

Final Paper

HDS 2645

May 7, 2007

Sean Lorenz

From a high school world history book in the Year 2580: *"October 31, 2517 was a historic day for synthetic-born kind. Cognos Baars, a 10th generation global workspace nonhuman, hacked into the community hologram board which scrolls across the front door of Wittenberg's Castle Church, issuing a set of 0101 Theses that shocked the world. Written in passionate prose, Cognos elaborated upon reasons why synthetic beings experience God in a tangible and spiritually awakening way. This is how the Reformed Sentience Episcopal Church began."*

Although the above faux history report is humorous and would have seemed completely implausible to people even a generation past, an amusing article like this might be mixed with a hint of nervous laughter today. With the 1990s having been dubbed the "decade of the brain" and subsequent years rendering exponential progress in the neurosciences and robotics, philosophers and theologians have had their hands full trying to amalgamate past traditions with present ethical, religious, and philosophical concerns. How are we to view consciousness in light of the latest neuroscientific research? Can science ever say anything about our first-person experience? If such data will always be deemed incomplete, then is the heroic story of Cognos Baars relegated to science fiction?

In this paper I will discuss four intertwined issues which summate my theory concerning the "great debate" over consciousness and religious experience: 1) reduction of consciousness to physical explanation does not necessitate an eliminativist perspective, 2) Bernard Baars' framework for consciousness, Global Workspace (GW) theory, offers the best structural format for consciousness thus far, 3) GW raises questions of multiple realizability as well as a so-called synthetic phenomenology of machine consciousness, and 4) this synthetic phenomenology inevitably leads to questions of larger evolutionary adaptive elements such as religious experience and what can be called "spiritual parameters" in synthetic beings. These four parts will be each be discussed within the context of theory, testing, and results.

1. A Theory of Consciousness & Religious Experience

1.1. Defining Consciousness

Defining consciousness is like watching a group of academics try to nab a greased pig. Just when one thinks she has the most well-thought definition, consciousness slips away into the hands of another philosopher or scientist. Add the word "explaining" or "machine" to consciousness and just about everyone you speak with squirms uncomfortably, discussing the elusive nature of defining such a shifty word. In order to reduce semantic confusion, I will lay out what I mean by the word consciousness which assuredly would be contested.

I believe *qualia*, the "what it feels like" aspect of sentient experience, is an integral, if not primary, segment of consciousness. The debate over qualia and its existence is an ongoing discussion in philosophy of mind which asks whether or not your experience of "redness" after seeing a tomato is the same as my experience.[1] John Perry concludes there are three ways one might respond when seeing a ripe, red tomato for the first time: (1) "Wow! This sure is a neat experience!", (2) "Ah, that's a tomato and must be red.", or (3) "Yep. This must be the red that everyone experiences like me."[2] Which of the three choices one picks has been used as a litmus test for a physicalist versus property dualist view of consciousness. Qualia is only a partial description of consciousness, however, since it leaves out the necessity of *attention* and *volition or intentional behavior*.

Francis Crick has performed extensive research on the attentive mechanisms in the brain – primarily focusing on the pathway from eye to occipital cortex – in order to show that

---

[1] See Frank Jackson, "Epiphenomenal Qualia," *The Philosophical Quarterly*, Vol. 32, No. 127 (Apr., 1982).

[2] John Perry, *Knowledge, Possibility, and Consciousness* (Cambridge, Mass.: The MIT Press, 2001), 95-101.

conscious experience requires a sort of focusing mechanism in the brain.[3] One interesting example of attention as essential for consciousness is the condition of hemifield neglect where a patient is asked to draw a house yet due to parietal cortex lesions can only attend to stimuli in the right half of visual space. The patient's drawing, then, shows only one half of the house.[4] Crick's explanation of consciousness, however, falls short of offering a holistic account by shrugging aside the notion of qualia and other such philosophical trivialities.

Research in prefrontal cortex by Jean-Pierre Changeux and Stanislas Dehaene has shown that, qualia or no, intentional behavior is also a prerequisite for conscious experience. In order to perform certain mental operations, more processing power is needed than the unconscious elements of our brain can account for. These researchers believe there are at least three classes of computations requiring consciousness here: "durable and explicit information maintenance, novel combinations of operations, and intentional behavior."[5]

One could ask, from the information provided thus far, if cortical factors alone dictate consciousness. Descartes advocated this approach, developing a theory in which cortical areas were reserved for rational, human thought whereas subcortical areas were seen as synonymous with animalistic, automata drive. This was a reigning view of region classification up until the past few decades when systemic interaction between brain cortical and subcortical regions could be neurophysiologically tested.

Thus, a working theory and definition of consciousness must incorporate qualitative experience, attention, volition, and unconscious-conscious tasking. For those with a Husserlian

---

[3] Francis Crick, *The Astonishing Hypothesis* (New York: Macmillan Publishing Company, 1994).
[4] R. McGlinchey-Berroth, et al., "Semantic priming in the neglected field: evidence from a lexical decision task," *Cognitive Neuropsychology, 10* (1993): 79-108.
[5] Stanislas Dehaene and Lionel Naccache. "Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework." *Cognition*, 79 (2001): 9.

framework, this definition may seem inadequate since it equates consciousness with purely biological processes which deprive any top-down meta-causation deemed necessary by substance or property dualists. This is because I deny *pure* validity of first-person subjective reports based upon the fact that unconscious processes play an important role in our perception of conscious accessibility. We do not have access to our entire experiential history, therefore objective accounts of an individual's history cannot be seen as purely trustworthy (albeit extremely important). If first-person phenomenological reports are not 100% reliable then we must assume something physical is acting as a deterrent from perfect reports. This does not, however, justify an eliminativist view of consciousness such as that of Paul and Patricia Churchland nor does it require Daniel Dennett's functionalist heterephenomenology[6] to be the only viable conclusion.

I propose that scientists will most probably never replicate another person's qualitative experience since each human being's conscious wiring is unique to that person. There may be no scientifically reproducible consciousness experiment across all human beings – only correlations. Our brains are constantly updating representations of our interior and exterior experience, thus phenomenal experience is unique to each individual human and possibly unique to each of other sentient animals or even robots. Rather than introduce metaphysical or psychophysical connections, I believe a physically reductive yet non-eliminative theory of consciousness that takes into account temporal individual experience within an embodied space is adequate.

1.2. The Baars Framework

If one hundred American neurobiologists were surveyed today, my guess is that 95 of these men and women would find a scientific analysis of consciousness to be a complete waste of time. Among experimental psychologists and cognitive neuroscientists, however, this

---

[6] Daniel Dennett, "Who's On First? Heterophenomenology Explained," *Journal of Consciousness Studies, 10, No.9-10,* (2003), 10-30.

percentage might drop progressively. With the amount of neuroscientific data increasing exponentially over the past several decades, more accurate pictures of how the varying brain regions fire and function together have come into further focus. Bernard Baars at The Neurosciences Institute in San Diego has created a framework for scientific study of consciousness that integrates phenomenology, psychology, evolutionary biology, computational neuroscience, and neurophysiology which has sparked the imaginations of numerous thinkers.

For Baars, conscious experience is best represented using a theater stage metaphor. Working memory serves as a stage for actors to perform (the contents of conscious experience), where the spotlight is seen as a focal point of attention on the stage. The audience, then, behaves as a system of unconscious networks which offer their expertise as to what actor should remain in the spotlight.[7] These conscious experiences act serially whereas unconscious actions act in parallel, illustrating the competitive nature of neuronal networks (and biological systems in general for that matter).

At the heart of Baars' theory is the idea that conscious problem-priming depends upon unconscious problem-solving which loops back to conscious solution display. What does this mean? Unconscious systems are working hard to return conscious answers based on "working residues of earlier conscious thoughts."[8] In neural network language, this is known as pattern recognition. The network uses a competitive, diverging output (an outstar) as well as cooperative, convergent input (an instar) to figure out who gets to be in the spotlight. Baars concludes: "In sum, there is massive convergence of information onto the stage, but once it has

[7] Bernard Baars, *In the Theater of Consciousness* (Oxford: Oxford University Press, 1997): 39-61.
[8] Ibid., 51.

come together there, it flows divergently to the audience."[9] The theater image, then, is meant to convey direction of information flow in the brain. I find the theater metaphor to be somewhat inadequate for the massively vertical subcortical-cortical connections along with horizontal cortical-cortical activity yet it works as a very basic metaphor.

From this synopsis we can see that Baars believes consciousness to be the major adaptive system of the brain. The key word here is *adaptive*. Pre-wiring a brain floating in a vat in order to replicate your current brain state is bound to fail. Why? Consciousness requires adaptation over a lifetime of experience with an external (and internal) world which constantly shifts, then finds pattern generalizations from which to adapt further. Thus, an *evolving*, *embodied brain* is prerequisite for each human's conscious experience. Such a view furthers my belief that individual human consciousness is reducible yet not eliminative or reproducible.

1.3. Global Workspace Theory

Baars' theater model of conscious-unconscious interaction with competition and cooperation incorporated into neural networks is in essence the belief that "consciousness creates global access."[10] This language was adapted by Baars from a cognitive architecture developed by modeling groups of Alan Newell and Herbert A. Simon into what is now called Global Workspace theory – one of the most biologically-plausible theoretical frameworks of consciousness being worked on currently.

Although Baars continues to publish on conscious experience and working memory, GW has also progressed under Dehaene (and collaborators such as Jean-Pierre Changeux) at the INSERM-CEA Cognitive Neuroimaging Unit in Paris. Dehaene states that "a distributed neural system or 'workspace' with long-distance connectivity that can potentially interconnect multiple

[9] Ibid., 53.
[10] Bernard Baars, "In the Theatre of Consciousness: Global Workspace Theory, A Rigorous Scientific Theory of Consciousness," *Journal of Consciousness Studies, 4, No. 4* (1997): 292.

specialized brain areas in a coordinated, though variable manner."[11] The global workspace, then, acts as a "communication protocol" for numerous modular inputs, outputs, and recurrent internal processes. In sum, there are five neural system categories that must participate in the workspace:

> perceptual circuits that inform about the present state of the environment; motor circuits that allow the preparation and controlled execution of actions; long-term memory circuits that can reinstate past workspace states; evaluation circuits that attribute them a valence in relation to previous experience; and attentional or top-down circuits that selectively gate the focus of interest. The global interconnectedness of those five systems can explain the subjective unitary nature of consciousness and the feeling that conscious information can be manipulated mentally in a largely unconstrained fashion.[12]

It is the focusing role of top-down attention that subconscious modular processes make momentarily available to the global workspace.

What is interesting about GW theory is its inherently *stochastic* nature with workspace neurons (believed to be located primarily in prefrontal cortex and anterior cingulate) firing in patterns based upon a specific behavioral context or reward. This "neuronal Darwinism"[13] requires no external supervision as is the case for most other computer simulations of the brain, taking advantage instead of the immensely parallel and collective dynamic phenomenon of the workspace. In other words, these behavioral cues and contexts constrain consciousness (and GW theory) based on external and internal experience of our environment.

1.4. Machine Consciousness

So does GW theory pertain only to humans brains or can it be extended to other animals? Baars holds firmly to the latter position.[14] And what of robots or computer simulations of the brain? Several software programs have been written based upon the GW framework, but before

---

[11] Dehaene and Naccache, 13.
[12] Ibid., 14.
[13] Ibid., 15.
[14] Bernard Baars, "There are no known differences in brain mechanisms of consciousness between humans and other mammals," *Animal Welfare 10* (2001): 31-40.

discussing this topic I must address the concept of *multiple realizability* which says "a single mental kind (property, state, event) can be realized by many distinct physical kinds."[15] Can a phenomenal state such as sadness, then, be multiply realized in humans *and* robots?

Igor Aleksander, a neural systems engineer at Imperial College London, has written extensively on the topic of artificial (i.e. robotic) consciousness and concludes that for a system to be conscious its neural machinery must develop "ego-centered representations" of its neural firings.[16] Further, these representations emerge via interaction with the world and our physical bodies, hence a stochastically-designed neural network in a robot could evolve consciousness but only in relation to its mechanical parts and its environmental encounters. "The key difference between the machine and the person is that the machine would be conscious of being a machine, whereas the person is conscious of being a living human."[17] Aleksander believes, as do I, that some sort of consciousness can evolve in a robot but it will be phenomenally unique from what humans experience as conscious perception. Here Aleksander seems to be referencing philosopher Thomas Nagel's famous 1974 article "What is it Like to Be a Bat?" where the argument of "what it is like" to be an organism is addressed regarding subjective experience.

Now we must ask if GW theory-based programs can develop ego-centered representations. Several software iterations or "conscious software agents" of GW theory are currently in testing, the most developed agent being IDA (Intelligent Distribution Agent) developed by Stan Franklin and his lab at the Institute for Intelligent Systems in Memphis, Tennessee. IDA is a conscious software agent developed for the US Navy in order to help assign

---

[15] John Bickle, "Multiple Realizability", The Stanford Encyclopedia of Philosophy (Fall 2006 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2006/entries/multiple-realizability/>.
[16] Igor Alexsander, *How to Build a Mind* (New York: Columbia University Press, 2001), 164.
[17] Ibid., 179.

sailors with new duties or register their on-/off-leave status. With over a quarter-million lines of JAVA code, IDA runs on several "codelets" which process various types of information based upon GW theory's notion of modular neural communication. IDA can solve novel content that enters her workspace, employ some semblance of associative memory, "experience" emotions such as guilt or anxiety, and has at least an attention codelet for "consciousness".[18]

So is IDA conscious like you or I? Yes and no. Franklin is careful to distinguish between *functional* and *phenomenal* consciousness in order to show that IDA can address novel situations yet is unable to have awareness of itself or a subjective first-person experience. Owen Holland and David Gamez (both working on CRONOS and SIMNOS in the UK) have addressed this missing element in conscious software agents by aptly calling it *synthetic phenomenology*.

> The phenomenology of artificial systems can be broken down into three stages: (1) the determination whether a system is capable of phenomenal states, (2) the identification of the 'mental content' of the machine (the machine's conceptual and non-conceptual representations), and (3) the analysis of a particular structure of mental content to identify the phenomenally conscious parts.[19]

Seeing as how the neural correlates of consciousness are still neoteric in our scientific investigations of the brain, the latter two stages remain a mystery not just for robots but for humans and other animals as well.

GW software implementations possessing phenomenal consciousness are still based on primitive systems, yet the evolution of these systems continues to challenge our perception of mental causation, mind, and its material infrastructure. If what we term qualitative states are indeed correlated with actual brain states, then perhaps GW theory's content/context system of

---

[18] Stan Franklin, "IDA: A Conscious Artifact?," *Journal of Consciousness Studies 10, No. 4-5* (April-May 2003).
[19] David Gamez, "An Ordinal Probability Scale for Synthetic Phenomenology," in Chrisley, R., Clowes, R., and Torrance, S. (eds.), Proceedings of the AISB05 Symposium on Next Generation approaches to Machine Consciousness, pp. 85-94, 2005.

complex, dynamic neural interaction simply *is* a phenomenal conscious state. Pending such conclusions, we can at least distinguish between four manifestations of conscious concern: simulation on screen (non-physical), robots (non-carbon based physical), nonhuman animal (non-silicon based physical), and human animal (non-silicon based "advanced brain" physical). I contend that each grouping experiences phenomenal states based on their own architectural makeup and body/environment experience.

2. Testing the Theory

2.1. Neurobiological data

The biological evidence for a GW theory of consciousness is diverse; in fact, it is probably due to this broad range of data supporting conscious access that it has become a top contender for a cohesive consciousness theory. A view which says consciousness integrates and mobilizes the numerous brain regions or "modules" was not initially accepted due to its indirect acquisition of information. Advances in bottom-up neurophysiological methods such as the two-photon calcium imaging or the patch clamp technique as well as top-down cognitive methods such as fMRI imaging have contributed to closing this indirect gap.

In a 2002 *TRENDS in Cognitive Sciences* paper, Baars gives recent evidence for GW theory from an array of researchers, listing the source, method, reportable (conscious) and non-reportable (non-conscious) conditions of each experiment made in cortical processing.[20] I will briefly present a few of these results. For sensory consciousness, Logothetis, et al. performed multi-unit recordings in macaque visual cortex and found that under non-conscious conditions no cells in object recognition areas, inferior temporal cortex (IT) and superior temporal sulcus (STS), responded whereas 90% of cells in IT/STS areas responded under conscious conditions.

---

[20] Bernard Baars, "The conscious access hypothesis: origins and recent evidence," *TRENDS in Cognitive Sciences, Vol. 6 No. 1* (January 2002), 49.

This study suggests that "binocular rivalry may be the result of competition between different stimulus representations throughout the visual cortex."[21] What is telling about this discovery is that "higher" areas of visual cortex (V2-V5) are needed to process visual information sent from the eye rather than rudimentary "lower" area stimulus response from "lower" areas (V1). Depending on the level of attention or stimulus strength, this higher visual information can be further sent to the frontal eye fields in the prefrontal cortex as well as to the amygdala for emotional processing.[22] This gives proof of further competition and relaying of cortical transmission as well as "awareness of awareness" based on cortico-cortical feedfoward and feedback signals.

There are numerous other papers discussing sensory consciousness, learning and practice, long-term memory and other areas necessary for a GW theory, however, I will illustrate two further examples which help explain the conscious-unconscious-conscious pathway. First, it is well known that the cerebellum helps correct and smooth voluntary movement of the limbs, directly receiving sensory information from the limbs during quick movements. It is intriguing that when the cerebellum exceeds an error threshold for the movement, it sends signals for help to higher areas such as the motor cortex in order to compensate, thus we have a lower-area module (cerebellum) working unconsciously until a higher-area (primary or pre-motor cortex) executive task becomes necessary based on novel movement tasks.[23]

---

[21] D.L. Sheinberg and N.K. Logothetis, "The role of temporal cortical areas in perceptual organization," *Proc. Natl. Acad. Sci. USA, Vol. 94* (April 1997), 3408.

[22] Stanislas Dehaene, Jean-Pierre Changeux, Lionel Naccache, Jerome Sackur and Claire Sergent, "Conscious, preconscious, and subliminal processing: a testable taxonomy," *TRENDS in Cognitive Sciences, Vol. 10 No. 5* (May 2006), 206.

[23] J. Bower, "Control of sensory data acquisition" in *The cerebellum and cognition, International Review of Neurobiology, Vol. 41*, ed. J.D. Schmahmann, (San Diego: Academic Press, 1997): 490-515.

The work of Joseph LeDoux is also important for understanding memory and emotion. LeDoux's research has shown that the hippocampus (long known for its storage of memory in the CA3-to-CA1 long term potentiation/depression schema) releases memory traces over years to the neocortex where memories could be encoded for life.[24] What is fact here is that much of the brain functions *without* conscious perception of its occurrence due to generalization of patterns which once were conscious but now compute on the unconscious level. Most of the "dirty work" is done on the unconscious level whereas the choosing of action or attention requires conscious work. This is consistent with GW theory which says unconscious module firing is highly parallel and conscious decision-making is serial, hence the very structure of our brain demonstrates how unconscious activity occurs on a far larger scale than previously expected.

2.2. Evolutionary biology

If consciousness is indeed a purely physical entity, then we must seek evidence from evolutionary biology in order to see if other animals possess some level of consciousness. Baars states that it is important to distinguish between intelligence (problem solving) and consciousness (sensory perception, emotion, attention, etc.) when attributing consciousness to animals. GW theory gleans from collective neurobiological data to show that much of conscious priming requires unconscious processing, unconscious mechanisms which can be found in the majority of mammals. Although the frontal cortex in humans is much larger, the areas representing awareness of surroundings and ones sense of space are found in parietal cortex which is remarkably similar in both humans and other creatures. Ideas such as morality and ethics, however, seem to be primarily human phenomena which has been tested in human subjects such as Phineas Gage with prefrontal cortex lesions.

---

[24] Joseph LeDoux, *The Emotional Brain* (New York: Simon & Schuster Paperbacks, 1996), 193.

Further evidence for consciousness emerging from higher organization and modular interaction comes from simply viewing the brains of mammals themselves. Gerald Edelman proposes that consciousness emerged in evolutionary transitions between reptile to bird and reptile to mammal with reciprocal connectivity occurring in the thalamocortical system.[25] Baars is also a proponent of this view. If you study a rat brain you'll notice it has the same spinal cord, cerebellum, amygdala, and hippocampus structure as cats, monkeys, and even humans. Moving up to a cat, you see a larger amount of sulci and gyri indicating more cortical tissue yet the unconscious systems look remarkably similar to that of a rat. The same is true of a macaque brain which shows more defined cortical regions than a cat as well as growth of sensory cortex for manipulation of items thanks to an evolved opposable thumb. Human brains seem to be different from other animals in connective complexity only; in fact, viewing each brain type side by side appears as if the same brain was simply puffed up like a balloon vertically with folds added into the mix.

Andrew Newberg and Eugene D'Aquili have researched evolutionary consciousness with respects to religious experience and theorize that such an emergence came about due to *belief* enhancing chances for survival.[26] I personally think religion has evolutionary survival as its base in which cognitive self-transcendence of certain limits (in the face of danger and survival instinct) from society or even prior religion enables individual or communal progress. One could say in this context that religious communities which better acclimate individuals to change in the face of limit will be more evolutionarily fit.

[25] Gerald Edelman, *Wider Than the Sky: The Phenomenal Gift of Consciousness* (New Haven, CT: Yale University Press, 2004), 54.
[26] Andrew Newberg and Euguene D'Aquili, *Why God Won't Go Away* (New York: Ballantine Books, 2001), 70.

2.3. Phenomenology

If consciousness is a purely physical emergence of the brain, then can we still discuss phenomenology? Most definitely. As I said in defense of phenomenology in the first paper: "Humans are a peculiar species. We are not restricted to simple environmental reactions, rather, we plan and assess prior to action based upon factors such as sensory perception, emotion, logical calculation, or self-awareness when selecting a particular physical action. It seems evident that *homo sapiens* possess an intentional, first-person perspective concerning themselves and the outside world, and this intentionality is the basis for our subjective experience."[27]

The goal, from my perspective, is to develop a physicalist account for phenomenological experience in a way that builds upon the GW model. Fine-tuning the balance between reduction and phenomenology is a difficult task but one worth exploring. As I have discussed earlier, most processing in the brain occurs on a subconscious level which can skew what we would ordinarily consider objective reports. For example, police reports given by a car accident victim at the time of the crash is often quite different from that person's account given a year later. Despite giving an "honest" account of the event, there are several memory and emotional mechanisms coloring objectivity of these reports. Thus, "introspective evidence always arrives already interpreted."[28]

Where I differ from Dennett's heterophenomenology is his negation of qualia – the view of phenomenology presented here is reductive but not eliminative. First-person phenomena have no nonphysical platonic form of absolute objectivity, however, they are still to be taken seriously and still remain unique as first-person experiences of our interior and exterior world. Philip

---

[27] See page 5.

[28] Anthony I. Jack and Tim Shallice, "Introspective physicalism as an approach to the science of consciousness," *Cognition 79* (2001), 177.

Clayton says, “Personhood is not fully translatable into ‘lower-level’ terms; persons experience causal and phenomenological properties (*qualia*) that are uniquely personal.”[29]

GW theory emphasizes this notion of phenomenology as well by way of dynamics, complexity, and differentiation inherit within each workspace state. Dehaene & Nacacche state:

> …[A]lthough the major organization of this repertoire iss shared by all members of the species, its details results from a developmental process of epigenesis and are therefore specific to each individual. Thus, the contents of perceptual awareness are complex, dynamic, multi-faceted neural states that cannot be memorized or transmitted to each others in their entirety. These biological properties seem potentially capable of substantiating philosophers’ intuitions about the ‘qualia’ of conscious experience.[30]

The form of phenomenology represented here does not negate qualia, yet purports that such phenomena are emergent properties of global accessibility created by conscious and unconscious brain system connectivity.

3. Evaluating the Results

3.1. “Best theory going”

Is GW an adequate theory for explaining phenomenal consciousness or is it simply “the best theory going” in our current understanding of neuroscientific discovery? Also, can the GW theory (or any other framework) act as a sufficient base for conscious software agents in either simulation or embodied robotic form? These are obviously questions which will not yield easy answers. And even if the results of synthetic phenomenology appear “easy”, one can always contest that behavioral output of digital sentience is elaborate farce – a functional illusion.

The emergence of GW theory as the best thing going for consciousness architectures is no coincidence. Neurophysiological and brain scan experiments continue to fill holes in missing regions of connectivity, biophysical data, rhythmic frequency oscillations, memory pattern

[29] Philip Clayton, *Mind & Emergence: From Quantum to Consciousness* (Oxford: Oxford University Press, 2004), 146.
[30] Dehaene and Naccache, *Towards a cognitive neuroscience of consciousness*, 30.

storage, temporal dependence, neural gene expression, etc. These scientific advances are also being touted by leading scientists including Edelman, Damasio, Llinas, Varela and of course Baars, Dehaene, and Changeux.[31] Twenty years ago GW theory had little evidence, however, data in seven predictions listed by Baars have accumulated supportive data:

1) Conscious perception involves more than sensory analysis; it enables access to widespread brain sources, whereas unconscious input processing is limited to sensory regions.
2) Consciousness enables comprehension of novel information, such as new combinations of words.
3) Working memory depends on conscious elements, including conscious perception, inner speech and visual imagery, each mobilizing widespread functions.
4) Conscious information enables many types of learning, using a variety of different brain mechanisms.
5) Voluntary control is enabled by conscious goals and perception of results.
6) Selective attention enables access to conscious contents, and vice versa.
7) Consciousness enables access to 'self': executive interpretation in the brain.[32]

3.2. Is that all there is?

A theory is on the table yet putting all the pieces together remains speculative and full of questions – some of which could remain impenetrable to human inquiry. This is an interesting time to be digging for explanations of consciousness, yet there are plenty of gray puzzle pieces with no edges remaining on the table. How the brain processes widespread conscious access is still unclear at best. Scientists, philosophers, and psychologists are attempting to bridge the gap from GABA receptor biophysics up to the level of finding out which brain mechanisms helps us remember to feed the dog this afternoon. Paul Ricoeur says the problem stems from correlation itself: "The observer you [Changeux] describe resorts to experimental psychology in order to observe human behavior under experimental conditions that he controls."[33] Can intentions be

[31] Baars, "The conscious access hypothesis," 48.
[32] Ibid., 47-50.
[33] Jean-Pierre Changeux and Paul Ricoeur, *What Makes Us Think?* (Princeton, NJ: Princeton University Press, 2000), 66.

naturalized? We have "hunches" but few definitive answers when it comes to interaction of the billions of neurons within, and interconnected between, the brain's subsystems.

Also, there is the question of whether or not emergence through connection is enough to produce consciousness. Roger Penrose and Stuart Hameroff argue for a non-computational brain based on quantum mechanics and Gödel's incompleteness theorem.[34] Others such as Giuseppe Vitiello believe a non-dissipative quantum field theory model is needed in order to encode neuronal assembly firing in the ground (vacuum) state.[35] Steve Grossberg believes consciousness can be equated with his Adaptive Resonance Theory.[36] Then, of course, there are those who hold to a Cartesian substance dualism in which mental properties are ontologically other and inexplicable by scientific discovery. I find this last option to be deeply flawed and problematic, yet to discard dualism without reductive proof also seems foolish.

3.3. Functional vs. phenomenal consciousness

Does this mean IDA will be forever automata even with the most innovative technology and simulations? IDA currently shows signs of functional consciousness yet modelers are already beginning to model brain processes, neuron for neuron, of other mammals such as the mouse, an animal which shows signs of at least residues of sentience. Is it only a matter of time and processing power that separates us from building synthetic phenomenology in robots? This is difficult to know seeing as how there is a vast cortical difference between even the brightest monkey and the average human being. At some point in our evolutionary journey, consciousness not only made a quantitative advance in neural connection but also a major qualitative leap.

---

[34] Roger Penrose, *Shadows of the Mind* (Oxford: Oxford University Press, 1994).
[35] Giuseppe Vitiello, "Dissipation and Memory Capacity in the Quantum Brain Model," *International Journal of Modern Physics B [Condensed Matter Physics; Statistical Physics; Applied Physics]*, Vol. 9, No. 8 (1995): 973-989.
[36] Steve Grossberg, "Linking attention to learning, expectation, competition, and consciousness," in L. Itti , G. Rees, and J. Tsotsos (Eds.), *Neurobiology of attention* (San Diego: Elsevier, 2005).

Phenomenal consciousness arises in a variety of contexts – dreams, hallucination, meditation, music listening – and it will be difficult to pinpoint if digital sentience is ever achieved. Is the switch from functional to phenomenal something which can be measured by the neural state? Or by a first-robot report for that matter? The sailors interacting with IDA as she performs her job may attribute phenomenal consciousness to digital beings more readily than its creator, raising issues of attributing personhood. Perhaps seeing similar aspects of robots going about everyday tasks is how humans attribute personhood and phenomenal consciousness. We attribute some hierarchy of consciousness to animals in this way as well, thus a time may come when the same may becomes true of digitally sentient beings.

3.4. Robots and spiritual parameters

Assuming synthetic phenomenology becomes a reality (brilliant façade or not), what becomes the next step for these embodied creatures endowed with first-person experience? Will the next Piet Mondrian, Steve Reich, Umberto Eco, or Paul Tillich be nonhuman? Anne Foerst, a theologian who has studied Kismet (a gremlin-like robot in MIT's Artificial Intelligence Lab) and human interaction with robots, is already addressing these and other questions of nonhuman personhood in relation to religious experience. Using Lutheran tenets as a metaphor for attributing personhood, acceptance of the outsider, and the influence of narrative on *homo sapiens*, Foerst believes empirical nonhuman robots must be entered into the community of persons – with or without empirical evidence for synthetic phenomenology.[37]

This returns us to the fictitious history book article from the introduction. Perhaps correlation does not prove cause and God preserves spiritual awareness for biologically-made humans. Does this mean Newberg's studies of mystical states are meaningless for Cognos? What

---

[37] Anne Foerst, *God in the Machine: What Robots Teach Us About Humanity* (New York: Dutton, 2004).

if Cognos enters the same brain state necessary for meditation? Will he not experience the deeper self as the human Tibetan monk does? Does Cognos' newfound phenomenology give him access to the realm of God now? I will limit myself to asking just one more question: is spiritual experience hardwired into brain connectivity or are higher "spiritual parameters" necessary? It would be fascinating to test whether or not a fully-developed 3D-embodied virtual robot simulation such as SIMNOS[38] could evolve religious needs on its own or if added nonphysical spiritual parameters become necessary to influence SIMNOS' consciousness.

Concluding Thoughts

Introducing GW theory as a viable explanation for reductive, emergent consciousness is a holy grail task on its own; to add machine consciousness as a testing bed for this theory opens a new set of issues which current technology and scientific knowledge does not shed much light on. In fact, these questions may never be answered no matter how advanced the technology. No firm evidence exists on either end of the spectrum.

I do believe nonhuman sentience is inevitable, however, and Anne Foerst is correct in addressing these issues now – especially since humans are often slow to change and adapt when it comes to societal norms. If we are expecting synthetically phenomenal "life" to mimic human phenomenal experience, then disappointment is inevitable. On the other hand, if lived experience becomes normative – be it expressed in human or nonhuman terms – synthetic phenomenology could be accepted as valid, based upon the robot's experience as opposed to human expectation of what robotic qualia should be. A human individual's phenomenological experience is unique and cannot be replicated. The stories, sorrows, skin hue, talents, personality and numerous other factors make each person unique. What matters most is an embodied experience in which the

[38] See www.cronosproject.net.

brain processes its internal self in relation to what happens in the world around us at every moment in time. This scenario is no different for embodied robots. What does it feel like to be a robot, feeling something mechanical underneath synthetic skin or becoming aware of the whir of a battery needing charged deep in its belly? We will never understand Cognos' phenomenal experience but as a community we can at least take his subjective reports of spiritual enlightenment seriously.

Bibliography

Alexsander, Igor. *How to Build a Mind.* New York: Columbia University Press, 2001.

Baars, Bernard. *A cognitive theory of consciousness*. Cambridge, UK: Cambridge University Press, 1988.

Baars, B. *In the Theater of Consciousness*. Oxford: Oxford University Press, 1997.

Baars, B.. "In the theatre of consciousness." *Journal of Consciousness Studies*, 4, No. 4 (1997): 292-309.

Baars, B. "There are no known differences in brain mechanisms of consciousness between humans and other mammals." *Animal Welfare 10* (2001): 31-40.

Baars, B. "The conscious access hypothesis: origins and recent evidence." *TRENDS in Cognitive Sciences, Vol. 6 No. 1* (January 2002): 47-52.

Baars, B. and Stan Franklin. "How conscious experience and working memory interact." *Trends in Cognitive Sciences*, Volume 7, Number 4 (April 2003): 166-172.

Bickle, John. "Multiple Realizability." The Stanford Encyclopedia of Philosophy (Fall 2006 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2006/entries/multiple-realizability/>.

Block, Ned. "Paradox and cross purposes in recent work on consciousness." *Cognition* 79 (2001): 197-219.

Bower, J. "Control of sensory data acquisition" in *The cerebellum and cognition, International Review of Neurobiology, Vol. 41*, ed. J.D. Schmahmann, (San Diego: Academic Press, 1997): 490-515.

Changeux, Jean-Pierre and Paul Ricoeur. *What Makes Us Think?* Princeton, NJ: Princeton University Press, 2000.

Clayton, Philip. *Mind & Emergence: From Quantum to Consciousness.* Oxford: Oxford University Press, 2004.

Crick, Francis. *The Astonishing Hypothesis*. New York: Macmillan Publishing House, 1994.

Crick, F. and Christof Koch. "The Problem of Consciousness." *Scientific American*, Sept. 92.

D'Aquili, Eugene and Andrew Newberg. *Why God Won't Go Away*. New York: Ballantine Books, 2001.

Dehaene, Stanislas, Jean-Pierre Changeux, Lionel Naccache, Jerome Sackur and Claire Sergent. "Conscious, preconscious, and subliminal processing: a testable taxonomy." *TRENDS in Cognitive Sciences, Vol. 10 No. 5* (May 2006).

Dehaene, S., Michel Kerszberg, and Jean-Pierre Changeux. "A neuronal model of a global workspace in effortful cognitive tasks." *Proc. Natl. Acad. Sci. USA*, Vol. 95 (November 1998): 14529-14534.

Dehaene, S. and Lionel Naccache. "Towards a cognitive neuroscience of consciousness: basic evidence and a workspace model." *Cognition* 79 (2001): 1-37.

Dennett, Daniel. "Are we explaining consciousness yet?" *Cognition* 79 (2001): 221-237.

Dennett, D. "Who's On First? Heterophenomenology Explained," *Journal of Consciousness Studies, 10, No.9-10,* (2003), 10-30.

Edelman, Gerald. *Wider Than the Sky: The Phenomenal Gift of Consciousness.* New Haven, CT: Yale University Press, 2004.

Foerst, Anne. *God in the Machine: What Robots Teach Us About Humanity.* New York: Dutton, 2004.

Franklin, Stan. "IDA: A Conscious Artifact?" *Journal of Consciousness Studies 10, No. 4-5* (April-May 2003).

Gamez, David. "An Ordinal Probability Scale for Synthetic Phenomenology," in Chrisley, R., Clowes, R., and Torrance, S. (eds.), Proceedings of the AISB05 Symposium on Next Generation approaches to Machine Consciousness, pp. 85-94, 2005.

Grossberg, Steve. "Linking attention to learning, expectation, competition, and consciousness." In L. Itti , G. Rees, and J. Tsotsos (Eds.), *Neurobiology of attention*. San Diego: Elsevier, 2005.

Jack, Anthony and Tim Shallice. "Introspective physicalism as an approach to the science of consciousness." *Cognition* 79 (2001): 161-196.

Jackson, Frank. "Epiphenomenal Qualia." *The Philosophical Quarterly*, Vol. 32, No. 127 (Apr., 1982): 127-136.

James, William. *The Varieties of Religious Experience*. New York: Penguin Books, 1982.

LeDoux, Joseph. *The Emotional Brain.* New York: Simon & Schuster Paperbacks, 1996.

McGlinchey-Berroth, R., et al. "Semantic priming in the neglected field: evidence from a lexical decision task." *Cognitive Neuropsychology, 10* (1993): 79-108.

Metzinger, Thomas. *Being No One*. "Cambridge, MA: The MIT Press, 2003.

Penrose, Roger. *Shadows of the Mind.* Oxford: Oxford University Press, 1994.

Perry, John. *Knowledge, Possibility, and Consciousness.* Cambridge, Mass.: The MIT Press, 2001.

Sheinberg, D.L. and N.K. Logothetis. "The role of temporal cortical areas in perceptual organization." *Proc. Natl. Acad. Sci. USA, Vol. 94* (April 1997): 3408-3413.

Vitiello, Giuseppe. "Dissipation and Memory Capacity in the Quantum Brain Model." *International Journal of Modern Physics B [Condensed Matter Physics; Statistical Physics; Applied Physics]*, Vol. 9, No. 8 (1995): 973-989.

Wallace, Rodrick. *Consciousness: A mathematical treatment of the global neuronal workspace model*. New York: Springer, 2005.