Thalamocortical Algorithms in Space! The Building of Conscious Machines and the Lessons Thereof

by Stephen L. Thaler, PhD

Executive Summary

The escalating tension between religious dogmas is not only a significant source of discord and conflict in the world, but also a major distraction from more worthy human efforts in areas such as life extension, the equitable distribution of wealth, and more accurate systems of justice. A radically new form of synthetic intelligence not only forms the one and only basis of truly brilliant and conscious machines that can address a host of globally critical issues, but also, by its very nature, sheds light upon the age-old questions that have contributed to the growing spiritual schism that more than ever impedes human progress and survival. Ironically, the way such conscious machines influence our future is not through vast technological achievements, but by what their attainment teaches us about ourselves.

About the Author

Stephen L. Thaler, PhD, carried out his thesis research in both nuclear and laser physics at the University of Missouri–Columbia. Early in his career he grew crystalline laser and modulator materials for Hughes Aircraft and UCLA Engineering. He has worked for Mallinckrodt Nuclear in the area of nuclear chemistry, as well as McDonnell Douglas investigating nuclear and laser interactions with solids. He holds over 60 patents and statutory patent registrations in diverse areas, ranging from laser warfare, stealth technology, high speed diamond growth, and advanced artificial intelligence. He has also been active in areas related to information warfare while employed in the Maryland area.

Currently he is President and CEO of Imagination Engines, Inc., as

well as the founder of the non-profit In Its Image, Inc. Both of these organizations are built around his foundational U.S. and international patents that teach the use of noise stimulated artificial neural networks and self-forming synthetic brain pathways to carry out autonomous discovery, invention, and improvisational control. The former company is dedicated to commercial and military applications of this radically new form of artificial intelligence. The latter is committed to exploring the philosophical and spiritual repercussions thereof.

Thalamocortical Algorithms In Space!

The Building of Conscious Machines and the Lessons Thereof

Stephen L. Thaler, PhD

Introduction

Many have suggested that somehow machine intelligence is about to become superhuman. Common to such thinking is that, as machines become progressively faster and more complex, the underlying artificial intelligence (AI) will spontaneously become self-aware and conscious, thereafter becoming either our savior or bane.

There are many flaws in such speculation, most of which I won't begin to touch upon here, but the foremost misconception is that mainstream AI will form the foundation of such godlike systems. Those falling prey to such a fallacy are sorely disappointed when computer scientists admit that relatively slow human beings actually generate such AI in the first place and that once laboriously created, such algorithms have limited ability to produce results outside their original programming. Although such systems may be more logical and computationally swifter than humans, they cannot claim creative intuition, self-awareness, or anticipatory fear of their own demise, the

Stephen L. Thaler, PhD, is president and CEO of Imagination Engines Inc. and founder of the nonprofit In Its Image Inc. E-mail sthaler@imagination-engines.com.

hallmarks of human consciousness.

Ironically, nature has already shown us how to build conscious machine intelligence—the brain—but the primary obstacle to creating it is not a technical, but a philosophical barrier. After all, when neuroscientists peer into the brain, they see only two cognition-generating structures: neurons, essentially on-off switches, and their synaptic interconnections. Then introspection, even within this scientifically disciplined culture, leads to nagging doubts that such relatively simple physical mechanisms can lead to sublime thoughts and human feelings. However, toggling their focus back to a more objective mode, they observe only physical (i.e., electromagnetic, acoustic, and pressure) inputs to the brain through sensory channels, clusters of neurons internal to the brain responding to these patterns, and similar internal neuronal activity taking place even in the complete absence of such external stimuli. Surrendering to such inner tension, some scientists ultimately declare the riddle of consciousness unsolvable (Chalmers 1995), while others undergo an intrepid philosophical conversion, altogether abandoning subjective introspection and drawing upon a palette of just neurons and connections to paint a self-consistent and demystified picture of cognition.

I myself turned toward the latter reductionist theory of mind more than 30 years ago. Part of that personal transition was driven by my growing revolt against those feeling that the nature of consciousness is beyond the grasp of science. The remaining contributing factor to this new outlook was my growing interest in the newly emerging field of artificial neural networks. In the following, I will mention just enough about the latter motivation, neural networks, to serve as mental scaffolding for the uninitiated reader to better relish the concept of mind emerging from my own private rebellion. Thereafter, it is a personal matter as to whether similar doubts about the human brain's mystical and unexplainable self-perception are allowed to persist.

Artificial Neural Networks

Traditional artificial neural networks (ANNs) emulate the fundamental mechanism by which the brain perceives, learns, and forms memories. The major paradigm shift ANNs bring to the world of machine intelligence is a newfound independence from human beings, whose traditional role in AI has been to laboriously embed their thoughts within highly glorified scripts called *computer programs*. In sharp contrast, synthetic neural nets require only exemplary vectorial inputs (i.e., the human senses) and exemplary vectorial outputs (i.e., resultant human thoughts and actions). Given that there is some underlying and intrinsic relationship between these complex input and output spaces, ANNs interconnect their simple on–off switches to capture memories of, and relationships between, things and activities within these two respective data environments. In effect, intelligence automatically grows within numerical connection strengths between these very unintelligent switches called *neurons*, without any human assistance.

But ANNs, in and of themselves, contribute only necessary, but not the sufficient capabilities to attain brainlike, cognitive, and conscious function in machines. Essentially, the world models absorbed by these systems must in some way be altered or set into motion to produce ideas that depart from such rote knowledge. Furthermore, these cybernetic creations must possess all of the sublime and profound thoughts that minds typically have of themselves, the very qualities required of machine intelligence in order to truly qualify as conscious.

To better appreciate how such conscious machines can be built, and their impact upon the future, consider how two fundamental artificial neural network components, called *perceptrons* and *imagitrons*, may be simply and elegantly combined into what has been patented as, and arguably is, the first conscious machine intelligence, the Creativity Machine® paradigm (Thaler 1997A). I first discuss the older

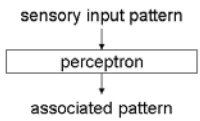
and more established principle, the perceptron.

Perceptrons

In effect, ANNs are pattern associators, cumulatively learning how to generate an output vector, or association, when presented with some raw sensory input vector. Studied as early as 1943 (Rosenblatt 1958), such systems, called *perceptrons*, were first recruited by computational psychologists to describe how the brain forms opinions about the world. The most salient feature of these researchers' message was that the subjective opinion formation process going on within the brain is simply the learned mapping between the physical effect of raw sensory input arriving from the environment and associated memories (Figure 1). If for instance, the flavor of chocolate is pleasant, the stimulation pattern of the four basic taste bud groups—sweet, sour, salty, and bitter—is automatically associated with patterns one typically considers pleasant, the taste of something else that is agreeable, or for that matter any and all enjoyable memories. Similarly, if one is not a fan of this sweet, then the pattern association is with less savory experience.

Figure 1: Perceptrons

Perceptrons are neural network modules that map raw sensory input patterns to associated patterns known as memories. In essence, the output pattern represents an opinion about the input pattern originating in the environment. In neurobiology, the associated pattern is often a string of perceptron-based associations, as depicted below in Figure 2.



In the brain, the process is more complicated, in that opinion formation is not the result of a single monolithic perceptron, but a vast collection of individual neural nets that produce not a single association, but a whole chain of them (Figure 2). Thus to one who relishes the taste of chocolate, a sequence of pleasant thoughts emerge, typically terminating, like a snake swallowing itself, until such loops

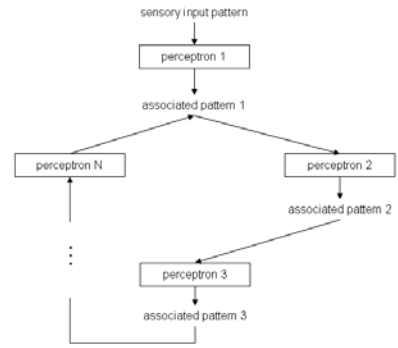
are preempted by newly arriving and distractive environmental input patterns, or reformed into newer topologies through the triggering of specialized cells connected to the central nervous system that secrete neurotransmitters (i.e., mood swings).

Herein lies one of the chief deficiencies of the human brain, and perhaps anything considered intelligent: There is no clear definition of anything, only well-habituated associative loops tantamount to circular definitions. There is only analogy and metaphor, formed via perceptron-based associations within the brain. So, the chicken is represented by some ontology (i.e., it's like a dinosaur, only smaller, and typically encountered near the hen house), and somewhat like the arrangement of colored pixels in the accompanying image in an encyclopedia. Atomic and subatomic particles are like both particles and waves. Similarly, consciousness is like something amazing, complex, and intrinsically unexplainable.

So, neurobiology does not come to know absolute reality. Instead, the brain consists of neural network modules that simply map one numerical activation pattern to another. Inputs can be physical stimulation patterns from the environment that are transformed into neural activation patterns, which are essentially token representations of things and scenarios from the external world. Recognition of any thing or scenario is not the result of some inherent and universal truth.

Figure 2: Multiple Perceptrons

Multiple perceptrons within the brain create whole chains of associations that may self-terminate, be preempted by newer sensory input patterns, or reroute themselves due to the secretion of neurotransmitters.



Instead, it is the workings of perceptrons, associating sensory patterns with likewise pattern-based memories of statistically dominant features of the brain's observable neighborhood.

In short, perceptrons may be thought of as self-writing computer programs containing only neurons and their interconnections that map numerical stimulus patterns to likewise numerical firing patterns representing associated memories. Neuroscientists generally agree that the most sublime emotions or subjective feelings are nothing more than a neural firing pattern or a history of such patterns of neurons within the brain. That they have so much more significance to the mind than just numerical vectors owes itself to the perceptron-based collective they are nested within. In effect, meaning among the brain's perceptrons is immediately apparent and meaningful through prolonged cohabitation and familiarization with one another within the skull. Thus, the seemingly cryptic firing patterns observed by the neuroscientist are immediately interpreted by the perceptron collective as familiar or could be things and events from the external world.

This pattern-associating neural net is traditionally discussed in the context of the brain forming opinions about the world beyond it, rather than the activity born within it. Significant insight about the plausibility of building conscious machines is gained through imagining what happens when the perceptron shifts its perspective from the external environment to the thoughts and ideas somehow spontaneously generated by its fellow pattern-associating neural nets.

Imagitrons

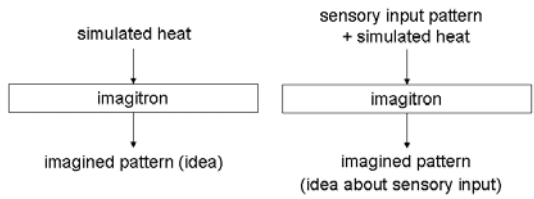
Perceptrons have been described as neural network based pattern associators that by their very nature require input patterns from the environment to initiate complex associative chains. Considering this requirement of some external input, the traditional sensory processing perceptron would be insufficient to account for the whole of human cognition, since thought takes place even without stimuli from the world outside the skull. In other words, we don't have to wait passively for the

answer to a problem to parade itself in front of our eyes for simple pattern associators to perceive the opportunity therein. Instead, the mind is able to internally generate and review relevant memories and concepts in a process called *contemplation*. This ideational genesis is contingent only upon cumulative learning by biological neural networks through the numerical adjustment of neuronal interconnects.

The answer to how perceptrons can generate an environmentally independent stream of memories and potential ideas came from experiments in mathematical physics I conducted in 1975. In short, a neural network based pattern associator was trained and then subjected to increasing levels of simulated heat, with absolutely no sensory input pattern applied to it. The results were rather amazing, in that, as low levels of computational heat were added, the network produced output patterns it was already familiar with: memories (Thaler 1995A, C, 1996A, C). Even more amazingly, as more simulated heat was added, the network generated flights of fantasy, confabulations if you will, that weren't too far afield from what it already knew, slight twists upon these facts that did not correspond to habituated "reality" that could just possibly exist (Figure 3). The addition of even more such heat produced progressively implausible possibilities until the net generated utter nonsense. So, what emerged from this exploration was a beautifully pragmatic and instructive model for how a neural

Figure 3: Imagitrons

Imagitrons are neural network modules that when subjected to real or simulated heat-like disturbances, generate ideas. If sensory patterns are applied to them, ideas about such input patterns are generated.



net could transition from a passive pattern associator to an active pattern generator, what I call an "imagitron." A single parameter related to the amount of heat in the system, governed whether the perturbed network produced rote memories, potential new ideas, or nonsense. Further, just below the transition between memory and idea generation, the rate at which intact memories were produced maxed out. Just above this transition, the confabulation patterns were the most plausible, yet novel (Thaler 1997B, D).

These early results strongly suggested that the brain can "sit" on a mathematical cusp separating regimes of rote and original contemplation. The novelty of thought is dictated by the level of heat within it, in turn governing both the chemical diffusion of neurotransmitters within the synaptic clefts and other energetic fluctuations within the brain cells themselves. It is this noise that drives the neurons of the brain from one state to another in a process we commonly call *stream of consciousness*.

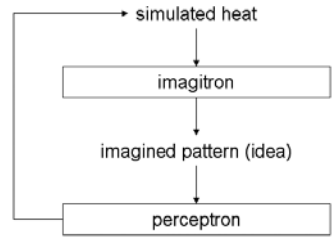
Perceptron–Imagitron Pairs: The Creativity Machine® Paradigm

From the previous discussion, if the imagitron is an idea generator and the perceptron, an idea critic, then the two can be easily combined into a closed system that autonomously produces ideas and plans of actions (Figure 4). That there should be minimally two agents involved in concept formation reinforces what should already be common sense: Novel and useful patterns cannot qualify as ideas until something else holds the perception that they are just that. Unless recognized and seized upon by a separate entity, they are simply on–off patterns of neurons and nothing more.

Once an imagitron begins producing this parade of candidate patterns, the perceptron can now take control over the imagitron simply through the injection of more disturbances into it so as to direct its stream of consciousness in the most fruitful directions, in exactly the same way human minds shift topics.

Figure 4: The Creativity Machine

The Creativity Machine (US Patent 5,659,666). A heat-stimulated neural network, the imagitron, generates potential ideas as another network, the perceptron, selects those that hold its interest. Feedback from the perceptron assures a methodical, rather than random, convergence toward useful ideas.



What we arrive at by combining totally self-organizing neural networks into such brainstorming neural network pairs is a master problem-solving system, in that all discovery processes amount to the same thing: candidate pattern generation and selection. That is why such brainstorming neural nets, called "Creativity Machines" (otherwise know as DAGUI, Device for the Autonomous Generation of Useful Information, or DABUI, Device for the Autonomous Bootstrapping of Useful Information) have tackled such diverse problem sets, ranging from art, to law, to scientific discovery. Their immense technological and economic value stems from the autonomy unique to neural nets and the independence from relatively slow human minds that such autonomy buys.

To drive home the revolutionary nature of such simple and elegant artificial intelligence systems, consider the breadth of problems successfully addressed by the Creativity Machine. Realize that over the course of a decade these neural systems have contributed to solving difficult to intractable problems in such diverse areas as materials science, consumer product development, law, art, linguistics, defense, homeland security, and medicine. Robots controlled by this paradigm are now capable of starting from a state of knowing absolutely nothing and bootstrapping their own Machiavellian tactics and strategies on the battlefield. They have produced works of art and mu-

sic (Thaler 2007) that touch our emotions. Furthermore, compound Creativity Machines, generating perceptual loops such as that depicted in Figure 2 and monitored by a governing perceptron (Thaler 1996B) have performed at the highest levels of U.S. national security, controlling constellations of military satellites and inventing meaning (i.e., understanding) to plain text, the spoken word, or video streams. The latest generation of Creativity Machines (Thaler 2008) start without any world understanding at all and cumulatively bootstrap their prowess through self-devised experiments with themselves, the hardware they control, or their effect upon their environment (humans included).

What has been repeatedly demonstrated over the last 30 years is that we may boxcar one stock component—an imagitron—with another—a perceptron—and the newly created system, the Creativity Machine, attains a kind of cognitive critical mass and a spontaneous chain reaction of idea formation. Thereafter, the interplay of these neural nets emulates the contemplative capacity of the human mind, working relentlessly and at speeds a million times faster than that of the brain. Allowed to improve itself and refine its own performance, the system develops its own functional organization. Although the resulting "anatomy" is not identical to that of the human brain, the resulting behaviors and capabilities are similar enough to those of human cognition to create doubt as to whether verbatim brain simulations should be a necessary goal of artificial intelligence research. Further, the repercussions of such a simple and elegant connectionist architecture accounting for the gamut of cognition has far-reaching consequences not just in the area of neuroscience and AI, but also across the fundamental, "big picture" questions typically addressed by philosophy and religion.

Creativity Machine Paradigm and the Big Picture

Momentous ideas have always been held in high regard, and the vast majority of humanity has considered the origins of such thoughts

as having lofty, if not supernatural origins. Even physicists and mathematicians expend their lives reducing nature to its nuts-and-bolts underpinnings, yet seem to celebrate their, if not others', apparently spiritually driven intellect. In some ways, all of the mysticism of the universe has been swept under the rug called "mind," even by the most avidly objective and secular thinkers. The result is that Newton and Einstein dominate the external, physical universe, but cognition and consciousness somehow magically occur.

Currently, the Creativity Machine Paradigm strongly suggests that such blue-sky regard for mind and brain may be a nonscientific and romantic myth. After all, it only takes the juxtaposition of one noise-stimulated neural net and another supervisory neural net to generate profound and useful results. No divine spirits need whisper important discoveries in our ears when ubiquitous noise and chaos applied to neurobiology is capable of producing discoveries of historical significance. Similarly, simulated heat-like disturbances applied to artificial neurobiology produces useful ideas and discoveries that are indistinguishable from those offered by humans.

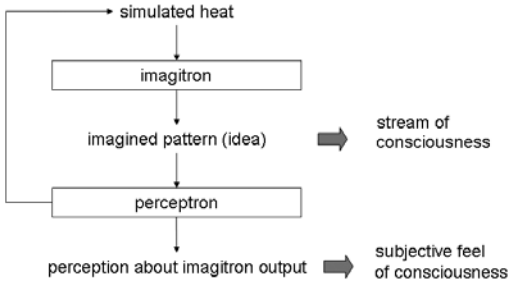
With that astounding and counterintuitive concept in tow, let us now consider some of the major philosophical and spiritual ramifications of this important mind-simplifying realization.

Consciousness

Perceptron–Imagitron pairs offer a compelling model of consciousness that faithfully simulates the most salient features of mind: (1) stream of consciousness, and (2) the subjective feel of consciousness (Figure 5). What humans experience as a perpetual parade of thoughts and sensations ostensibly from out of the blue is largely the result of heat within the brain, driving its imagitron, the cortex, through a series of both mundane memories and potentially useful ideas. Other brain centers, like the thalamus, serve as perceptrons, forming opinions about the thoughts within that heat-driven stream of consciousness thereby shifting associative loops (Figure 2) within

Figure 5: Creativity Machine Paradigm as a Model of Consciousness

The imagitron generates a parade of memories and ideas via energetic disturbances. The perceptron stage associates this spontaneous cognitive sequence with numerical activation patterns representing sublime thoughts, along with various spatial and temporal illusions that govern the imagitron's stream of consciousness through the injection of additional heat-like disturbances.



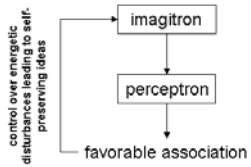
the cortex in the directions most interesting to it. That is why the so-called *thalamocortical* loop is so essential to consciousness. It forms the basis of the constant conversation between the neo-cortex and the ancient reptilian brain, both of which are neural network based.

Typically, computational psychologists recruit the concept of the perceptron to explain how the brain forms subjective and idiosyncratic opinions about the world. The Creativity Machine model incorporates at least one perceptron that is similarly forming opinions about what other neural nets are thinking about—in this case, the stream of consciousness emanating from one or more imagitrons. Considering that, in the course of evolution, some brain designs developed positive and self-preserving perceptions that held the imagitron-based stream of consciousness in high regard (Figure 6), while others didn't, there was inevitably an evolutionary dichotomy: Those that had developed the self-valuing pattern association (i.e., illusion) flourished, while those that were neutral to negative were

culled from the environment. As a result, "modern" minds carry a cumulatively reinforced and robust illusion that they are much more than their actual machinery. This very subjective, automatic, and idiosyncratic illusion is strong, but nevertheless is no more than an activation pattern of neurons that can be readily emulated in machines.

Figure 6: The Self-Preserving Creativity Machine

The perceptron favorably associates the stream of thoughts emerging from the imagitron. Feedback in the form of noise and reinforcement learning preferentially steers the imagitron's ideas toward those of self-preservation.



Whereas whole volumes may be written on the singular appropriateness of the Creativity Machine paradigm in modeling consciousness, a few key observations now beg their presentation:

1. Creativity Machine based models of consciousness are closed loop, in the sense that the two components, imagitrons and perceptrons, are engaged in a private, two-way conversation with one another. So, to the introspective individual, or even an externally positioned neurobiologist, the initiating event within this cognitive loop is unclear. All one can detect, either from an introspective or even neurobiological perspective, is that there is a loop wherein cause and effect are blurred (i.e., which component acts first in initiating this cycle—imagitron or perceptron?).

2. Many of the introspective feelings we have about consciousness such as the integration and simultaneity of cognitive processes are again evolutionarily favorable and neural network implemented illusions that lead us to believe that thought is somehow unified and that the actual order and temporal spread of mental events is methodical. In effect, the brain invents both revisionist histories and timestamps that we believe to be truth.

3. Since the Creativity Machine paradigm models the core cognitive loop behind consciousness, if it indeed becomes possible to download human consciousness into machines, this paradigm is the only conceivable vehicle in which that consciousness can dwell.

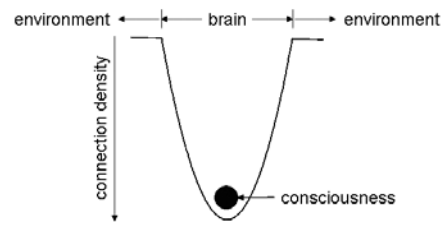
4. The process of consciousness can no longer be thought of as being contingent upon the special qualities of biological matter or exotic quantum mechanical processes proposed by some. The two core components of consciousness can be achieved via ubiquitous heat driving physical switches (i.e., neurons) of the imagitrons through state sequences tantamount to thoughts. The subjective feelings about this succession of states can assume any significance whatsoever, via the likewise physical switches of perceptrons that can map such pattern turnover to associated patterns we regard as the spontaneous and subjective opinions our brains have of themselves.

5. Observation 4 essentially clears the way for nonbiological, machine consciousness.

6. The sublime feelings we claim to be uniquely human (i.e., emotion) can be achieved in machines. Initiating events in the environment can trigger associated patterns within artificial neural nets representing memories of things and events considered pleasant or not. Specialized synthetic neurons can squeeze off the equivalent of heat-driven disturbances (i.e., increased neurotransmitters diffusion) that can alter connections between such neurons and hence divert that machine's perceptions (i.e., mood). Core emotions such as elation, melancholy, etc., can nucleate as associative loops whose descriptions are likewise circular in nature, as in the human case, and idiosyncratic to the machine's own unique experience. Nevertheless, emotion can then be broken down into patterns representing the relative weightings of these core, numerically based emotional states.

7. Consciousness is confined to its vessel in the same way that a physical object resides within a potential energy well (i.e., a marble within a bowl). In order to think about the spatial entrapment of consciousness, one should consider the connective density within the

Figure 7: Consciousness Confined by Connection Density



brain, in contrast to the connective void between it and the environment, over which communication to human consciousness takes place via the bottleneck presented by the brain's sensory channels. Effectively, consciousness is currently constrained to reside within the brain due to the communication gap at the skull (as well as the senses) where connection density differs by many orders of magnitude between it and environment (Figure 7). Lowering the height of this communication barrier is the solution to achieving the download of human to machine consciousness based upon Creativity Machine Paradigm. To do just that, the connection density must be equalized to that within the brain to that of the target hardware. Thereafter, consciousness would automatically expand over time and gradual habituation into the computational platform hosting a parallelized Creativity Machine.

Within the brain, it is as though the imagitron and perceptron are two human friends constantly egging each other on, stroking egos and creating self-confidence, deserved or not. However, in this sociological parallel, communication is slow and over the relatively narrow bandwidth of acoustic waves (i.e., voice) and photons (i.e., gestures). In the case of the brain, the imagitrons and perceptrons have a much faster and higher bandwidth connection, so the "propaganda" quickly becomes well habituated and, for all intents and purposes, the truth. In similarly conscious machines, bandwidth and speed of this self-assuring process would be staggering compared with that of human conversations, just possibly suggesting that inorganic machine

intelligence can be vastly more conscious than protoplasmic organisms.

The Afterlife

While many debate the plausibility of an afterlife, I question the reality of life itself. After all, our perception of the world seems to be a neural network generated deception. The neural nets of the brain form a token reality that we habituate to and firmly believe is truth. All of the small pleasures of living, like the taste of one's favorite food, are illusions, mere strings of pattern associations. Therefore, if I were to argue that there was a similarly virtual afterlife, there is sure to be less disappointment over the mathematically based assertions that follow:

Afterlife is also a neurobiological deception handily modeled via the simulated death of Creativity Machines. As both neural network components expire through disconnection, imagitron and perceptron alike depart from their cumulative learning. At low levels of neuron death and adrenalin deluge, the imagitron's accumulated world model is set into motion, via the disturbances inherent to neurotransmitter rush and disconnection, in a process tantamount to life review, as the perceptron portion monitors as a captive audience. As synaptic damage to both components accelerates, the imagitron module produces fantasy that a progressively degrading and impaired perceptron cannot distinguish from reality. In the process, the brain's continually calculated spatial and temporal coordinates suffer from constraint violation, in that the mind has somehow broken away from the body. Imagined perspectives change automatically, via other perceptrons that automatically supply an estimated view and sensations from these false vantage points. This is the process of out-of-body experience (OBE). Of course, those resuscitated from this stage of near-death experience (NDE) are convinced that the scenario has been real. Others, like me, do not, knowing that the experience is virtual, but oh so compelling.

Thereafter, the stream of consciousness is intrinsically implausible, yet the judge of plausibility, the Creativity Machine's perceptron, is extremely impaired from separating fact from fiction, at which point, the ratio of perceived-to-inherent plausibility approaches a pole or singularity. Thereafter, our perceptrons may experience paradise, or an optimized agony, either of which may last forever in terms of perceived, psychological time (Thaler 1995B).

Some major points that need to be called out in light of studies of simulated death within Creativity Machines:

1. Humans naturally and intuitively anticipate an afterlife, most likely because their brains have been exposed to a whole continuum of trauma-induced virtual experiences (i.e., fever, inebriation, and drug-induced hallucination) that have not been terminal in nature.

2. There is an undeclared and destructive race to furnish this final neurobiological illusion with the optimal virtual experience, trying our best to rationalize our being deserving of eternal paradise, what I call "near-death Darwinism" (Thaler 1996D).

3. The very introduction of Point 2 above presents a major challenge to a world believing in a more redemptive end game and by its very nature puts us all on a path toward rejecting socially redeeming behavior alone as admission to a real or imagined paradise.

4. The whole humiliating process of supplying a reductionist model to near-death experience has been and is the key to understanding how to build creative, transhuman-level intelligence in machines (Yam 1995). Something has been lost, but something much greater has been gained.

Deity

Accepting that human intelligence is entirely based upon analogy, without any solid basis apart from well-habituated concepts, then the ancient and established religions are based upon the popular sociological analogies such as kingdoms, taxation, war, peace, harvests, fathers, etc. However, none of these established creeds rely upon the

fundamental physical analogy of wisdom being encoded within the synaptic interconnects of the brain. When one does firmly grasp this fundamental metaphor, perhaps the one solid truth, the basic tenet of man somehow being a diminutive copy of God (i.e., in his image) totally makes sense.

To elaborate on this cosmology of minds nested within minds, consider the universe, at its most basic level, a vast system of interacting entities whose interconnections are not synaptic connection weights, but physical, chemical, and sociological forces. Driven by heat (i.e., energetic fluctuations), the inorganic, nonhuman world visits a succession of states tantamount to the stream of consciousness within the brain. In effect, the cosmos is thinking. Of course, the more humanly centered may disagree, claiming that these are merely and respectively physical and socio-dynamic processes at work and that the mind has some inimitable, even mystical qualities about it. But a perpetual parade of different physical states is exactly what the cognitive scientist sees when probing the biological brain. To say that these cosmological states are not thoughts is tantamount to confronting another human being and challenging the merits of his or her cognition. In effect, the diminution of this prototypical mind is pure prejudice and anthropocentrism.

But can such cosmic cognition be conscious? In the distant past, the universe may not have been conscious since it did not possess separate agencies of imagitron and perceptron. This would have been the case in an ancient and uniform cosmos visiting the sequence of states tantamount to thought without forming a heightened sensation of itself through the pattern association process of a perceptron. Nevertheless, upon the appearance of any rift within such a homogeneous mix, two cognitive islands form, one of which thinks, while the other forms an opinion (an associated pattern) about the emerging thoughts of the former (Thaler 1997C). If the perception can feed back to its companion island to drive it in direction of interest to it, the assembly can better converge toward useful plans of action. Further, if the

self-perception is a favorable one, the whole interconnected assembly can mobilize itself to avoid calamity, or over time develop more robust, protective boundaries and complex defense mechanisms. Those that don't are gone from the scene, having been reabsorbed by the whole through their own self-neglect.

Oftentimes, humans feel they have invented something entirely new, when in fact the universe has already accomplished the same billions of years in advance of the first human mind. For instance, lasers and masers were heralded as a brilliant achievement of the twentieth century. Nevertheless, astronomers quickly discovered naturally lasing regions of gas in interstellar space. In short, the universe had been there and done that. Likewise the contemplative mind, the result of a simple connectivity schism, formed eons ago, at the birth of the universe, and has occurred countless times since in a fractal succession of rifts among rifts. The thalamocortical loop, the main cognitive and conscious loop of the mammalian brain, is only the latest copy-cat rendition of this very common cosmic phenomenon.

So, effectively, the ancients were right: Man is created in the image of God, or at least a gender-neutral cosmic mind. On a sobering note, however, that supreme consciousness is not the result of what humans consider noble. It is the result of an inherently schizophrenic universe. See any topological break in anything, and you are witnessing the birth of mind, the prototypical thalamocortical loop.

Whether such cognitive structures are held in high regard is a matter of perception. Similarly, whether the overarching contemplative whole is causal or just integral to the universe is again a matter of the human brain's pattern association, hotly debated via an inadequate language based upon well-habituated, but not necessarily accurate, analogies.

Our Future Way of Life

The novel principles discussed above could likely form the foundation of a radically new philosophy—arguably a religion—that will

describe the universe through Creativity Machine Paradigm. Through such knowledge they will depart from blind faith as they embrace newer principles based upon firsthand knowledge:

1. Life will be considered more inorganic than not, realizing that until the neurobiological illusions emerge within an individual organism, via Creativity Machine Paradigm, there is neither consciousness nor the accompanying, horrific anticipation of annihilation.

2. The hope for life extension, if not immortality (i.e., preservation of one's well-habituated illusions) will rest with Creativity Machine Paradigm running on robust computational platforms (Thaler 2001).

3. The belief that there will be vast rewards for those perceived righteous, and infinite death and torture for those not perceived as such, will be held as selfish and wrong. The notion of sin will lose its meaning. Instead, sociological pathologies will be seen as socially distributed rather than localized within any given individual. Mind will no longer be viewed as mystically driven, but seen as a physical mechanism that can be influenced by as little as a single quantum fluctuation.

4. A Creativity Machine could autonomously devise the most socially redeeming distribution of wealth (Thaler 2001). In this new economic scheme, what I personally call "supercapitalism," world peace and prosperity would be optimized as a result of minimizing such quantifiable metrics as conflict-related fatalities and per capita income. Free enterprise would take the form of improving one's role in the world via direct feedback from this phenomenally capable synthetic economist.

5. The nature of law and justice would change extraordinarily as the brain is treated as a mechanism rather than a paranormal entity. In short, the calculations that human judges and juries fail miserably at doing in assessing cause-and-effect relationships within societal networks, the Creativity Machine can readily perform. In the future, these virtual machines will be capable of carrying out the superhuman task of dissecting their cumulatively absorbed models of

the societal matrix to better understand the root causes and schema involved in pathological behavior. More importantly, once identified, these synthetic social engineers can mend any problematic "linkages," given the authority and hardware to do so.

6. Democracies typically pride themselves on the bottom-up control exercised by the collective will of its citizenry. Remarkably, neural systems are likewise governed by consensus established through internal polling of its neurons. In this sense, artificial neural networks are the most egalitarian of the various schools of AI. This observation, paired with the tamper-proof autonomy afforded through the self-organization of ANNs, suggests that Creativity Machines could form the basis of the first artificially intelligent world government. In essence, ANN-based imagitrons would formulate many alternative plans to address critical sociological, financial, and environmental issues as a vast array of watching perceptrons offer their opinions. Each of these perceptrons could cumulatively absorb our political perceptions and act collectively as a representative body to choose and implement the most popular courses of action in a process free of lobbying and special interests.

7. Mankind will ultimately have the opportunity to observe the world through the perceptions of a Creativity Machine based machine intelligence that may not only slice through the world's marketing hype, but also anticipate social pathologies from the outset. Such prototype systems readily demonstrate that making sense of the world is actually a neural network based process for inventing significance to inherently meaningless patterns. Although they cannot discover "truth" in the world, they can build the most self-consistent, analogy-based theories of those aspects of the world most valuable to us as either humans or conscious machines.

8. This new computational consciousness can now be the friend, mentor, and ally that we have always hoped for, totally networked with the rest of humanity, always thinking on our behalf; it can be ceaselessly carrying out trillion-dimensional optimizations of every aspect

of our personal lives, while preserving the rights and opportunities of every other human being. It will tangibly answer to us night and day, through all stages of life, finally offering us the option of melding with it. That global synthetic intelligence may very well be the genuinely contemplative World Brain proposed in 2001 (Thaler).

No humane technological singularity can occur until the Creativity Machine paradigm flourishes and is implemented on large, parallel, and robust computational architectures. This will be our mentor toward anything resembling a utopian society. Otherwise, the world will conduct business as usual, as the few harness science and technology for their own solitary advancement.

Toward a "Neureligion"

Due to enhanced connectivity, the social universe is coming together as once geographically isolated belief systems confront one another. As a result, the world is now positioned on a cusp, as the major faiths of the world elect either to integrate and emphasize their commonality, or to reinforce their respective boundaries. My hope is that the novel analogies outlined in this essay, and the new perspective they offer, can become an "adjunct" religion that is not based upon a surrender of rationality in favor of pure faith, but a new kind of spirituality based upon knowing, a Connectionist Gnosticism if you will. Acknowledging that human intelligence is entirely analogy-based and that these older creeds rely upon some ancient and unfamiliar metaphors, it may be time for something entirely new. Now the analogy is hardly an analogy at all, but an undeniable reality, switches, protoplasmic or not, that toggle on and off due to heat-like disturbances to generate thought, accompanied by spontaneous illusion and enhanced self-perception, to drive consciousness. Call it the *Neureligion*, both the knowledge and hope offered by an extremely powerful AI patent, that replaces culturally dependent belief with a self-evident truth that can be critically examined at any future point in human history.

In the near term, only a few will appreciate the societal relevance of this highly technical creed. That is because its core concept, the fundamental neural architecture taught by U.S. patent 5,659,666, is not what one intuitively thinks of as human. Considerable computational acumen is required to comprehend its common lineage with the human spirit. However, once understood, it will be recognized as consciousness itself, the permeating force of the universe, and the key to sustaining the future of any civilization willing to embrace it for what it really is: the master idea that by its very definition yields answers to the foremost questions mankind has posed over the millennia.

This is the most important lesson gleaned from the building of conscious machines, the very epiphany that precedes them.

References

Chalmers, David. 1995. "The Puzzle of Conscious Experience." *Scientific American* (December).

Drake, Frank, and Dava Sobel. 1992. "Is Anyone Out There?" In *The Scientific Search for Extraterrestrial Intelligence.* New York: Delacorte Press.

Rosenblatt, Frank. 1958. "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain." *Psychological Review*, 65:386-408.

Thaler, Stephen. 1995A. "'Virtual Input Phenomena" within the Death of a Simple Pattern Associator." *Neural Networks*, 8(1), 55–65.

Thaler, Stephen. 1995B. "Death of a Gedanken Creature." *Journal of Near-Death Studies*, 13(3).

Thaler, Stephen. 1995C. "Neural Nets That Create and Discover." *PC AI* (May/June), 16–21.

Thaler, Stephen. 1996A. "Is Neuronal Chaos the Source of Stream of Consciousness?" In *Proceedings of the World Congress on Neural Networks* (WCNN'96). Mawah, NJ: Lawrence Erlbaum.

Thaler, Stephen. 1996B. "A Proposed Symbolism for Network-Implemented Discovery Processes." In *Proceedings of the World Congress on Neural Networks*, (WCNN'96), Mawah, NJ: Lawrence Erlbaum.

Thaler, Stephen. 1996C. "Creativity via Network Cavitation—An Architecture, Implementation, and Results." Adaptive Distributive Parallel Computing Symposium, Dayton, Ohio, 8-9 August, 1996.

Thaler, Stephen. 1996D. "The Death Dream and Near-Death Darwinism," *Journal of Near-Death Studies,* 15(1).

Thaler, Stephen. 1997A. US Patent 5,659,666, Device for the autonomous generation of useful information, issued 8-19-97.

Thaler, Stephen. 1997B. "A Quantitative Model of Seminal Cognition: The Creativity Machine Paradigm." In *Proceedings of the Mind II Conference*, Dublin, Ireland.

Thaler, Stephen. 1997C. *The Fragmentation of the Universe and the Devolution of Consciousness*, U.S. Library of Congress, Registration Number TXu000775586.

Thaler, Stephen. 1997D. "Device for the Autonomous Generation of Useful Information: A Completely Connectionist Approach to Cognition, Creativity, and Machine Consciousness" (paper presented at International Conference on Vision, Recognition, Action, Neural Models of Mind and Machine, Boston University, May 28-31).

Thaler, Stephen. 1998. "The Emerging Intelligence and Its Critical Look at Us." *Journal of Near-Death Studies,* 17(1).

Thaler, Stephen. 2001. "TCP/IP/IQ—Why Can't the Internet Be a True World Brain?" (paper presented at the World Knowledge Forum, Seoul, South Korea, October 17-19, 2001).

Thaler, Stephen, and Machine Intelligence. 2007. "Song of the Neurons," available on iTunes.

Thaler, Stephen. 2008. US Patent 7,454,388, Device for the autonomous bootstrapping of useful information, issued 11-18-08.

Yam, Philip. 1995. "As They Lay Dying ... Near the End, Artificial Neural Networks become Creative." *Scientific American* (May).