

UNPACKING NEOSENTIENCE: DYSTOPIAN TECHNO-EVOLUTION?

13
01

2009

by BILL SEAMAN

Contemporary and historical literature surrounding the creation of intelligent machines is vast and full of strong differing opinions. Lovelace, in her Notes by The Translator (as cited in Babbage, 1961) imagined a creative machine with the notion that machines might come to compose music and/or explore different kinds of ‘operational’ processes . McCulloch and Pitts formulation of the artificial neuron in the early 1940’s (McCulloch & Pitts, 1965) sparked the birth of a new field, where human bio-functionality could potentially be abstracted in the service of creation of machines. Turing’s writing on the potential of situated intelligent machines with “input” and “output” organs (Turing, 1986); his test for machine intelligence; his early articulation of the potentials of the field, in Computing Machines and Intelligence (Turing, 1990) are all central. Among other things John von Neumann compiled the first draft on the EDVAC...He adopted the McCulloch and Pitts symbolism in diagramming the logical structure of the proposed computer and introduced terms such as organ, neuron, memory...(Dyson, 1997) Artificial Intelligence was coined in a conference at Dartmouth in 1956 by John McCarthy. In 1958 John McCarthy and Marvin Minsky founded the Artificial Intelligence Laboratory at MIT. Minsky wrote many books on the subject. Society of Mind (Minsky, 1986) discusses the notion of ‘Agents’ — microprocesses that are unintelligent in themselves but are emergent in nature when interacting, enabling “intelligence” to arise. Minsky’s more recent writings concern machine emotion. (Minsky, 2006) Seaman and Rössler write that Neosentient entities may have programmed “force field” drives or surrogate emotions, in part, informing their potential interactive behavior with people and other machines.

Can autonomous intelligent machines be created and what will the nature of their phenomenology be? How will this differ from human phenomenology? Certainly, a phenomenology that arises out of embodied machine sensing will create a knowledge of the world that is ‘of itself.’ Yet, the cybernetic bonding of humans with machines complicate the difference between machinic sensing and technologically extended human sensing.

Ross Ashby's *Design for a Brain*, tackled many problems surrounding the creation of a situated thinking machine and adaptation. (Ashby, 1952) McCorduck, in *Machines Who Think* (McCorduck, 1979), a rich compendium of ideas surrounding the origins of AI, quotes Ross Ashby and then points to his concept of self-organization:

"The free living organism and its environment, taken together, form an absolute system... the two parts act and re-act on one another." (Ashby, 1952) This notion is not new, not with Ashby or even Wiener, for Ashby quotes scientists as early as 1906 who made the same observations. But Ashby refines it, introducing other concepts such as stability, a mode of survival in the organism... A key passage focuses this idea: "A determinate 'machine' changes from a form that produces chaotic, un-adaptive behavior to a form in which the parts are so coordinated that the whole is stable, acting to maintain certain variables within certain limits - how can this happen?" The answer is that the machine is a self-organizing system that responds to stimuli, changing its behavior and in some sense its shape, in order to achieve stability - what Ashby chose to call ultra-stability.

The goal for some (Seaman and Rössler in particular) is to work toward the creation of a synthetic self-organizing techno-species. Although many approach this goal in a positive light, "survival of the fittest" in the Darwinian sense becomes one key to the fear of intelligent machines. Will intelligent machines take over, replace and/or control people? To some extent this can already be witnessed in terms of the field of robotics, where machines have replaced many factory workers, and AI systems have replaced particular kinds of analysts e.g. should your loan be granted? Expert AI systems are used to help diagnose particular diseases, replacing potential analysis by doctors.

Ray Kurzweil has articulated an excellent Chronology outlining the flow of discovery related to the Age of Intelligent Machines. His most recent book, *The Singularity is Near* (Kurzweil, 2005), discusses in depth his thoughts surrounding the emergent change that intelligent machines might bring about. He states "In the 1950's John von Neumann, the legendary information theorist, was quoted as saying that 'The ever accelerating progress of technology...gives the impression of approaching some essential singularity in the history of the race beyond which human affairs , as we know them, could not continue'" (Kurzweil, 2005), Kurzweil states that "...A serious assessment of the history of technology reveals that technological change is exponential. Exponential growth is the feature of any evolutionary process, of which technology is a primary example." (Kurzweil, 2005), In defining his notion of Singularity, Kurzweil presents the following quote from Vernor Vinge's book, *The Technological Singularity*:

When greater-than-human intelligence drives progress, that progress will be much more rapid. In fact, there seems no reason why progress itself would not involve the creation of still more intelligent entities— on a still shorter time scale. The best analogy that I see is with the evolutionary past: Animals can adapt to problems and make inventions, but often no faster than natural selection can do its work— the world acts as its own simulator in the case of natural selection. We humans have the ability to internalize the world and conduct “what if’s” in our heads; we can solve many problems thousands of times faster than natural selection. Now, by creating the means to execute those simulations at much higher speeds, we are entering a regime as radically different from our human past as we humans are from the lower animals. From the human point of view, this change will be a throwing away of all of the previous rules, perhaps in the blink of an eye, an exponential runaway beyond any hope of control. (as cited in Kurzweil, 2005)

This is where we make a jump — when “intelligent” robotic simulations intermingle with physical environments and generate actual situated behaviors. Historically, human beings have sought to be in control of machines, not the other way around. As machines become autonomous, this notion will have many grey areas concerning personhood, slavery, human/machine interaction and notions of ‘cultural difference’.

Kurzweil has spoken of both positive and negative aspects of such change. Bill Joy presented a strong negative argument concerning technological discovery in his text *Why The Future Doesn’t Need Us — Our most powerful 21st-century Technologies - Robotics, Genetic Engineering, and Nanotech - Are Threatening to Make Humans an Endangered Species*. In particular Joy points out the danger of systems that can self-replicate, as well as potentially “spawn whole new classes of accidents and abuses”.

In counter distinction to Joy’s dystopian vision, Rössler (Theoretical Biologist and Physicist) and Seaman, (Artist/Researcher) have been researching the potential of generating an intelligent, situated, multi-modal sensing, computer/robotic system that would be benevolent in nature. Two differing approaches include the creation of such a system via the embodiment and integration of a series of conceptual methodologies developed by Rössler and Seaman (Seaman and Rossler 2007) utilizing a parallel processing computational system, aptly entitled *The Benevolence Engine* (Seaman and Rossler 2007); the second methodology seeks to posit a new paradigm for computing through the generation of an *Electrochemical Computer*, a multi-modal sensing system and related robotic environment — *The Thoughtbody Environment*. (Seaman and Rossler 2007); Both approaches are deeply informed by bio-mimetics and bio-abstraction. Much earlier, von

Neumann stated: “A new essentially logical theory is called for in order to understand high-complication automata and, in particular, the central nervous system. It may be, however that this process logic will have to undergo a pseudomorphosis to neurology to a much greater extent than the reverse.” (as cited in Dyson 1997)

We consider a Neosentient computer to be a system that exhibits the following functionalities: It can learn; intelligently navigate; interact via natural language; generate simulation potentials before acting in physical space; be creative in some manner; come to have a deep situated knowledge of context through multi-modal sensing apparatus and integrated software systems; and It displays mirror knowledge. The above work is scientific in nature and draws from multiple research domains including Artificial Intelligence, Artificial Life (in that the systems will first be examined and explored through computer-based emulation), Cognitive Science, Theoretical Biology, Engineering, Psychology, Robotics and the Arts. Speaking of an Electrochemical Computer I am thinking first about human beings and how they function as sentient bio-mechanisms. Although computers are often compared to brains, the mind/brain functions in a very different manner to that of the computer. In the context of generating artworks surrounding the Benevolence Engine and Thoughtbody Environment, as well as researching it in terms of speculative inquiry, I am approaching this question both as a branch of scientific research and as a form of conceptual art.

Informed and inspired by the ongoing research dialogue with Rössler and others, Seaman has been creating a series of Artworks/installations — A Video Tape with an extensive poetic text by Seaman – The Thoughtbody Environment / Toward A Model for an Electrochemical Computer; A series of Photo/Text images; A set of short Haiku-like techno/poetic texts — The Thoughtbody Interface, and the development of a proposal for a relational multi-modal database to house both the scientific research surrounding this project as well as aspects of the poetic work. A new work entitled Communication<->Space is in progress for Center Nabi, Korea.

Central to the project of generating a Neosentient robotic system is to better come to know ourselves — the qualities of being human, thus the project is paradoxical in that one attempts to make a Neosentient entity and in so doing better comes to know the human. When one seeks to employ a series of “living analogies” (Seaman and Rössler 2007) to abstract from the human the salient aspects of their nature, one is urged to address the subtle qualities of being alive as well as being self aware.

ART HISTORICAL AND CULTURAL PRECURSORS

There is a long history and mythology surrounding the creation of intelligent entities reaching back to Pygmalion and Prometheus. The invention of intelligent machines has also at times been shown in a “hostile” light in literature, across the arts, and within scientific discourse. Thus, an un-accepting world potentially becomes a “hostile” or an “extreme environment” for the arising of new forms of synthetic cognition.

One might ask how does this set of works and research agenda fit into the history of art and literature? Perhaps the most intriguing question relates to the notion of creating a work of art that can come to speculate on itself in an informed manner. One might provide the myth of Pygmalion as a starting point, although the research is not about constructing the “ideal woman” but a form of Neosentient entity. As the story goes, Aphrodite brought Pygmalion’s sculpture to life.

... Pygmalion concentrated on his art until one day he ran across a large, flawless piece of ivory and decided to carve a beautiful woman from it. When he had finished the statue, Pygmalion found it so lovely and the image of his ideal woman that he clothed the figure and adorned her in jewels. He gave the statue a name: Galatea, sleeping love. He found himself obsessed with his ideal woman so he went to the temple of Aphrodite to ask forgiveness for all the years he had shunned her and beg for a wife who would be as perfect as his statue. Aphrodite was curious so she visited the studio of the sculptor while he was away and was charmed by his creation. Galatea was the image of herself. Being flattered, Aphrodite brought the statue to life. When Pygmalion returned to his home, he found Galatea alive, and humbled himself at her feet. Pygmalion and Galatea were wed, and Pygmalion never forgot to thank Aphrodite for the gift she had given him. He and Galatea brought gifts to her temple throughout their life and Aphrodite blessed them with happiness and love in return.

Sally Everding traces online a number of related works that stem from the Pygmalion Myth – books, plays, paintings, poems, movies, writings related to AI etc. The book (she mentions) by Richard Powers, *Galatea 2.2* (Powers 1995) is an interesting example. This story explores the relationship between a human and the neural net system he is training to become knowledgeable in comparative literature. The exploration of the potentials of an intelligent computer’s relation to its human counterpart in this book illuminates both the positive and negative aspects of emergent learning systems. The correlation here is that systems can also learn and adapt to negative behavior.

We are not trying to make an aesthetically driven robotic artwork. We are however attempting to articulate a model for a Neosentient mechanism.

Alternately, the notion of machinic sexuality is particularly uncanny. How should a neosentient entity look? This issue has been broadly covered in science fiction. Perhaps, “Rachel” as a generated bio-entity that is sentient in *Blade Runner* is the most famous. Although here one begins to find a blur between bio-entities that are genetically engineered and electrochemical machines of the future.

The movie *AI*, 2001 directed by Spielberg presents a highly realistic AI boy. The concept of The Uncanny Valley addresses both a fascination and repulsion with differing levels of anthropomorphic abstraction.

The Uncanny Valley is a hypothesis about robotics concerning the emotional response of humans to robots and other non-human entities. It was introduced by Japanese roboticist Masahiro Mori in 1970, although drawing heavily on Ernst Jentsch’s concept of “the uncanny,”... Mori’s hypothesis states that as a robot is made more humanlike in its appearance and motion, the emotional response from a human being to the robot will become increasingly positive and empathic, until a point is reached beyond which the response quickly becomes that of strong repulsion. However, as the appearance and motion continue to become less distinguishable from a human beings, the emotional response becomes positive once more and approaches human-to-human empathy levels. This area of repulsive response aroused by a robot with appearance and motion between a “barely-human” and “fully human” entity is called the Uncanny Valley. The name captures the idea that a robot which is “almost human” will seem overly “strange” to a human being and thus will fail to evoke the empathetic response required for productive human-robot interaction.

This human attraction/repulsion mechanism to intelligent machines may cause great problems in terms of human/robotic relationships of the future. Kurzweil, in *The Singularity is Near*, states that “Strong AI promises to continue the exponential gains of human civilization... but the dangers it presents are also profound precisely because of its amplification of human intelligence.” (Kurzweil, 2005)

The dystopian story of *Frankenstein* by Mary Shelley (Shelley 2000), also subtitled *The Modern Prometheus*, was originally published in 1818. In a text by Ed Friedlander MD, *Enjoying “Prometheus Bound”*, by Aeschylus, Friedlander states: “Ovid names Prometheus as the god who made humankind in godlike form from clay, and says that maybe the creative power of the era gave us intelligence.” The up-dating of this myth in *Frankenstein* becomes another dystopian precursor. *Frankenstein’s* “Monster” started out as a benevolent creature but later turned violent in relation to the behavior bestowed upon him by his human counterparts. We are certainly deeply aware of the ethical issues surrounding our research and also the

“monstrous” potentials that it carries with it. Yet, we see our project, the creation of a model for a Neosentient entity, as having positive human values and find the ethics surrounding discussions of the project to be central as juxtaposed to the research into intelligent military machines and related systems — the deliberate construction of Killbots. There are a series of popular movies that explore the dystopian theme of machine control — where robots and/or sentient entities take power and seek to rule the earth — War of the Worlds; Terminator; West World; The Matrix trilogy; 2001 A Space Odyssey; and many more. These works play on the human fears that surround machine intelligence and robotic power. Alternately (and deeply frightening to my sensibility), the military are currently developing ‘smart’ weapons— robotic fighters, automated drone aircraft, and unmanned destructive vehicles. Again, we are taking on this research both with a deep ethical caring and a humanist agenda which must be seen as a critique on the negative potentials of such entities.

The relationship between humans and neosentient machines is a potentially charged one. One perspective might be to look at notions of cultural difference. A new machinic culture would be ‘of itself’. Neosentient machines would have a phenomenology that relates to the potentials of their acute sensing systems, their ability to share and transmit knowledge, to navigate and interact with their human counterparts and other Neosentient machines and/or Human/Neosentient hybrids. The potential direct transmission of knowledge from one machine to another can be seen as being both positive and negative. Teilhard de Chardin in his fascinating book, *The Phenomenon of Man* discusses the “Omega Point,” an evolutionary movement toward a unified consciousness. He states: “The point here is that this ‘something’—construction of matter or construction of beauty, systems of thought or systems of action—ends up always by translating itself into an augmentation of consciousness, and consciousness in its turn, as we now know, is nothing less than the substance and heart of life in process of evolution.” (de Chardin, 1959) Other philosopher/researchers like Roy Ascott and Pierre Levy as well as George Dyson, author of *Darwin Among the Machines* (Dyson 1997), have also written about the potentials of particular forms of technological connectivity in terms of a related unified sphere of consciousness. In the preface to *Darwin Among the Machines - The Evolution of Global Intelligence*, Dyson asks “Do we remain one species, or diverge into many? Do we remain many minds, or merge into one?” (Dyson 1997) Could the networking potential of new technologies and in particular Neosentient machines, enable an evolutionary shift to a new form of Neosentient cognition arising in a unified manner? Inversely, to what degree might the ability to transmit knowledge directly, generate a form of machinic schizophrenia? One might also ask, to what degree might

such a unified approach to consciousness have negative ramifications related to control — generating an intellectual panopticism. (Scharff & Dusek, 2003)

A humanistic approach in Japan is the potential creation of Robotic caregivers. As society ages, the need for expert caregivers can not be underestimated. Yet, cultural difference predicates both a deep love and deep repulsion of such entities in differing cultures. Perhaps the Animistic and Shintoistic heritage of the Japanese play into this particular cultural difference. “Astroboy” an exported Japanese robotic action hero also forms a playful “positive” precursor.

Another precursor to consider is the Golem myth, where a particular animated being is created from inanimate matter. The most famous Golem legend centered around Rabbi Löw, of 16th-century Prague. After molding the golem and endowing it with life, Rabbi Löw was forced to destroy the clay creature after it ran amok.

Gary Lochman discusses another precursor, the homunculus:

Prior to the rise of science and the mechanical vision of human life and the universe, the idea of creating human simulacra had a strong organic foundation. The homunculus was something one grew; the popular belief was that homunculi could be grown from the mandrake root, whose shape lent itself to anthropomorphic speculation.

As artificial life merges with its real counterpart, bio-engineering the definition of life itself becomes challenged as machines are figuratively “brought to life.”

So it is a fear of an entity’s emergent properties that can not be controlled, that might be central to our dystopian historical oeuvre. Thus, we have a fear of a potentially self-replicating emergent system that is of itself culturally, and perhaps exhibits the potential to become immortal. Alternately, the positive ethics surrounding the creation of a Benevolence Engine, and the discussion that it promotes, helps to balance such dystopian perspectives.

Special thanks to Otto Rössler and Jim Davies for discussion.

Bibliography

Ashby, W. R. (1952) Design for a Brain. New York: Wiley.

Babbage, C. (1961) Charles Babbage and his Calculating Engines: Selected Writings by Charles Babbage and Others. New York: Dover Publications, Inc. Written to clarify the work Sketch Of the Analytical Engine Invented by

Charles Babbage, by L. F. Menabrea

de Chardin, T. (1959) *The Phenomenon of Man*, translated Bernard Wall, Ltd. London: Wm. Collins Sons + Co. p. 178

Dyson, G. (1997) *Darwin Among the Machines*, The evolution of Global Intelligence, New York: Addison–Wesley Publishing

Kurzweil, R. (2005) *The Singularity is Near*, New York: Viking

McCorduck, P (1979) *Machines Who Think*. San Francisco: W.H. Freeman and Company. pp.82–83

Minsky, M. (1986) *Society of Mind*, New York: Simon and Shuster

Seaman, B. (2004) *Pattern Flows: Notes Toward a Model for an Electrochemical Computer — The Thoughtbody Environment*, delivered at the Cyberart

Bilbao, Conference, proceedings (forthcoming) also available at billseaman.com under Texts

Seaman, B and Rössler, O. (2007) *A Network of Living Analogies, Emoção Art.ficial 3.0, Interface Cibernética / organizacao Itaulab – Itau Cultural Center, Sao Paulo*, also found on billseaman.com

Seaman, B. (2005) *Pattern Flows | Hybrid Accretive Processes Informing Identity Construction*, *Convergence Magazine*, winter

Powers, R. (1995) *Galatea 2.2*. New York : Farrar, Straus, Giroux

Scharff, R.and Dusek, V. (2003) *Panopticism by Foucault in Philosophy of Technology : the Technological Condition : an anthology / edited by*. Malden, MA : Blackwell Publishers

Turing, A. (1986) Volume 10 in *The Charles Babbage Institute Reprint Series for The History Of Computing: A. M. Turing's ACE Report of 1946 and other papers*. Cambridge: MIT Press p.36

Turing, A. (1990) *Computing machines and Intelligence found in Mind design II : philosophy, psychology, artificial intelligence / edited by John Haugeland*, MIT Press, Cambridge, 1997. See also Turing, *Mechanical intelligence / edited by D.C. Ince*, New York.

McCulloch, W.S. and Pitts, W.H. “A Logical Calculus of the Ideas Immanent in Nervous Activity” in McCulloch, W.S. (1965) *Emodiments of Mind*. MIT Press.

Shelley, M. (2000) *Frankenstein*. Boston : Bedford/St. Martin's