

## **Artificial Human Nature + Warren Sack \***

“If design can be considered "the conception and planning of the artificial" then its scope and boundaries are intimately entwined with our understanding of the artificial's limits.”

Victor Margolin<sup>1</sup>

### **Artificial Intelligence as Design**

Artificial intelligence (AI) critics repeatedly ask whether humans can be replaced by machines: Can “human nature” be duplicated by machines and, if so, are humans then just a special sort of machine? By examining the present and history of AI criticism it is possible to identify moments where specific critics have fixated on particular qualities as the “essential” qualities of “human nature.” Reason, perception, emotion, and the body are four qualities that have been championed by AI critics (and proponents) as “essential” and (un)implementable as hardware or software machinery. I will argue that AI criticism’s preoccupation with the identification of “essentials” of “human nature” has left it blind – or at least short-sighted – to the cultural and ethical implications of AI: the ways in which AI technologies can influence the scope and boundaries of “human nature.”

I began by citing Victor Margolin and Richard Buchanan's definition of design because this essay addresses some of the points Margolin raises in his article entitled "The Politics of the Artificial." Margolin points out that AI scientist Herbert "Simon has gone so far as to call design a new 'science of the artificial.'"<sup>2</sup> Naturally, Simon places AI at the center of these new sciences.<sup>3</sup> Margolin critiques the artificial/natural binary assumed by Simon and argues that it is exactly at the border of these two terms that the work of design (and design criticism) is to be done. “When

---

+ Appears in the journal *Design Issues*, Volume 13, (Summer 1997): 55-64.

\* MIT Media Laboratory, 20 Ames Street, E15-120c, Cambridge, MA 02139, USA; tel: +617/253-4564; fax: +617/258-6264; email: wsack@media.mit.edu; <http://www.media.mit.edu/~wsack>

<sup>1</sup> Victor Margolin, “The Politics of the Artificial.” *Leonardo* 28 (5 1995): 349-356.

<sup>2</sup> *Ibid.*, page 349.

Simon compared the artificial to the natural he posited the natural as an uncontested term, ..."<sup>4</sup> This, as Margolin points out, is a highly problematic position for a designer of new technologies, like Simon, to attempt to defend. "As artificial beings like cyborgs or replicants more closely represent what we have always thought a human is, we are hard pressed to define the difference between us and them."<sup>5</sup> Thus, if a designer of new (AI) technologies assumes that "human nature" is "an uncontested term," the designer will be unable to articulate the ways in which new technologies could influence people's sense of self and others.

Margolin proposes that design work should be aimed at complementing (rather than replacing) the "natural," and, furthermore, that design should be based on a commitment to the spiritual: "A meta-narrative of spirituality can help designers resist techno-rhetoric that sanctions the continuous colonization of the natural."<sup>6</sup> He sees his work as an attempt to manage the borders between the artificial and the natural.<sup>7</sup> Margolin's work is, consequently, a reinvigorated modernist, humanistic project that sees humans as bound together – not by a set of uncontested, eternal "essentials" as an older, humanism might contend – but rather by a shared commitment, a shared "meta-narrative of spirituality" that should be continually defended by designers against the "artificial's incursion into the natural domain of our lives."<sup>8</sup> I do not agree with Margolin's position, but I think his argument offers a means of rethinking much of AI criticism because he shows how the modernist, humanistic assumptions of many AI critics and AI practitioners are contestable.

In this essay I will critique a number of AI practitioners and critics in a manner analogous to the way in which Margolin critiques Simon. Like Simon, many other AI practitioners and critics have assumed (and many still seem to assume) that the boundaries between "the artificial" and "human nature" are uncontested. Consequently, much of AI research and criticism has gone into

---

<sup>3</sup> Herbert Simon, *The Sciences of the Artificial. Second ed.*, Cambridge, MA: MIT Press, 1981.

<sup>4</sup> Op. Cit., Margolin, page 350.

<sup>5</sup> Ibid., page 354.

<sup>6</sup> Ibid.

<sup>7</sup> Ibid.

<sup>8</sup> Ibid., page 355.

investigating what is “essential” about humans.<sup>9</sup> I will argue that such AI investigations into “human nature” are oftentimes reminiscent of older, modernist, humanistic, philosophical debates about the same subject. For example, the philosophy of rationalism has been cited in critiques of “reasoning” in AI, phenomenology in critiques of AI’s conceptions of “perception” and “the body,” and romanticism has been named in defense of “human emotion” over machines’ supposed inability to experience emotion. By examining a series of these humanistic critiques of AI, it is possible to see how AI criticism has largely been a recapitulation, in miniature, of old, existing debates between rationalists, empiricists, romanticists, phenomenologists, pragmatists, and a handful of other named positions in the discourse of western philosophy.

A more rigorous critique of AI could, for instance, follow Margolin’s lead in accepting the artificial/natural divide encoded into the very name of the field (artificial intelligence) or -- in opposition to Margolin’s position -- could emphasize the permeability of the artificial/natural divide. The latter possibility, repeatedly cited by Margolin, is argued by the feminist theorist Donna Haraway, among others. Haraway proposes the cyborg as an ironic, utopic myth of human subjectivity.<sup>10</sup> By pointing out how inextricably coupled we all are to technology she provides a new vision of “human nature” based not on eternal, universal “essentials,” but on our shared fate as eclectic assemblages of a myriad of heterogeneous parts.

Design and design criticism is imbricated in a set of on-going processes to produce artifacts that address human needs, wants, and desires. Therefore, questions and assumptions of who “we” are as “humans,” what is/are our “nature(s),” and what constitutes the “artificial” or “artifactual” must be central to any discussion of design. I certainly will not argue, as someone like Herbert Simon might, that an understanding of AI is central to an understanding of design; but,

---

<sup>9</sup> Some AI projects have been attempts to design "intelligent" machines that do not, necessarily, exhibit “human-like” intelligence (e.g., the long history of AI, chess-playing programs). However, even these projects assume that the phenomenon of “human intelligence” is “natural intelligence” and thus identifiable *a priori* to any technological or “artificial” development. Thus, boundaries between “artificial” and “natural” phenomenon are assumed to be uncontested. Such was the case, for instance, with Edward Feigenbaum and others at Stanford developing "expert systems" for the field of chemistry: "Even though we set out to build a device whose behavior didn't have to match human behavior in detail, we ourselves didn't have any particularly good ideas about better ways of doing it, in ways different from the way, roughly speaking, human chemists do the same job." (Feigenbaum quoted in Pamela McCorduck, *Machines Who Think*. San Francisco: W.H. Freeman and Company, 1979, page 283).

I do think that AI criticism's questions and assumptions of these issues could be of interest to a wider audience of designers and design critics. "The Politics of the Artificial" are pivotal in larger political questions of design. An understanding of "human nature" as incontestable, universal, and timeless can lead to an Enlightenment politics of universal "brotherhood," but also underpins a politics blind to technological change. Furthermore, the sorts of modernist, humanistic philosophies assumed in much of AI discourse makes it ideologically impossible to state, from within the discourse, how the production and invention of AI technologies might change people and thus constitute an implicit political and ethical position at odds with the liberal humanism that it depends upon. It is, consequently, interesting to watch where the issues of ethics and politics resurface in a discourse like AI criticism.

### **Artificial Intelligence as Evil**

I intend to "map out" a variety of positions in the discourse of AI. At the center of my "map" is a recurrent moral outrage that erupts from the pens of disillusioned AI practitioners and other critics. This moral outrage can be found in older, well-known critiques, but the same sentiments are current today. Thus, I want to turn to a recent controversy published last year in the pages of *Interactions*, a journal for digital designers. The controversy concerns the AI technology of "digital agents."

Design "criticism" exploring digital agents (i.e., "autonomous" computer programs) has, predominantly, been a utopic discourse gushing descriptions of near-future pleasures and increased work efficiencies that will be provided by agents that can automatically find, filter, and summarize large bodies of information like the entire Library of Congress or the sum total of the World Wide Web pages posted on the Internet. In a publication for digital designers, *Interactions*, Jaron Lanier (one of the "fathers" of virtual reality) recently published a polemic against agents:

---

<sup>10</sup> Donna Haraway, *Simians, Cyborgs, and Women*. New York: Routledge, 1991.

“I find myself holding a passionate opinion that almost nobody in the ‘Wired-style’ community [*Wired* is the self-described “magazine of the digerati”] agrees with ... Here is the opinion: that the idea of ‘intelligent agents’ is both wrong and evil.”<sup>11</sup>

In a nutshell, Lanier’s rationale can be summarized as follows: “Agents” are a design discourse in which users and designers ascribe anthropomorphic agency to computer programs and so “The person starts to think of the computer as being like a person” <sup>12</sup> and “As a consequence of unavoidable psychological algebra, the person starts to think of himself [sic] as being like a computer.”<sup>13</sup>

### **Conflicting Humanisms**

By explicitly stating “I am ultimately arguing the merits of the humanist fantasy versus the materialist fantasy”<sup>14</sup> Lanier correctly locates his polemic as modernist, humanist position. But neither the term "modernist" nor "humanist" is precise enough to locate Lanier's position within the larger constellation of AI criticism. Of the "'Wired-style' community", Lanier explicitly names Nicholas Negroponte, a *Wired* magazine columnist, and director of the MIT Media Laboratory. Negroponte frequently extols the future virtues of AI agents<sup>15</sup> but also claims to be a humanist: “What needs to be articulated, regardless of the format of the man-machine relationship, is the goal of *humanism* [emphasis added] through machines.”<sup>16</sup>

If Lanier claims Negroponte as his opponent and yet also claims that his argument is "ultimately" one of humanist versus materialist, what is at stake if Negroponte too claims to be advocating a humanist position? I think Lanier has misrecognized his opponents as anti-humanistic

---

<sup>11</sup> Jaron Lanier, “Agents of Alienation.” *Interactions* (July 1995), page 66.

<sup>12</sup> Ibid., page 68.

<sup>13</sup> Ibid.

<sup>14</sup> Ibid., page 70.

<sup>15</sup> See, for example, Nicholas Negroponte, *Being Digital*. New York: Knopf, 1995.

<sup>16</sup> Nicholas Negroponte quoted in Stewart Brand, *The Media Lab: Inventing the Future at MIT*. New York: Viking, 1987, page 251.

materialists when, in fact, they are simply a different sort of humanist than the kind preferred by Lanier.

“Ultimately there is nothing more important to us than our definition of what a person is. ... This definition drives our ethics, because it determines what is enough like us to deserve our empathy. ... The artificial intelligence question is the abortion question of the computer world. What was once a research topic has become a controversy where practical decisions must reflect a fundamental ontological definition about what a person is and is not, and there is no middle ground.”<sup>17</sup>

In his ultimatums, Lanier simultaneously depicts the "abortion question of the computer world" as a black-and-white issue but also invokes a philosophical vocabulary (specifically, the vocabulary of "ontology") that implies more shades of gray can be found than his rhetorics might allow. In order to gain a better appreciation of the ethical stakes of AI debates (like the one which involves Lanier and Negroponte), I follow Lanier's lead into the territories of philosophy to show how Lanier and Negroponte are at odds with one another and are yet still both humanists.

Sherry Turkle (a sociologist, clinical psychologist, and well-known AI critic/advocate) calls positions, like the one occupied by Lanier, "romantic."<sup>18</sup> While Turkle seems to employ this term in its vernacular sense (i.e., as denoting someone who is impractical), I think this label can also be productively read as a name for one who is engaged with a type of thinking typified by the philosophical and artistic movement of the late-eighteenth century; namely, *romanticism*.<sup>19</sup>

---

<sup>17</sup> Op. Cit., Lanier, page 72.

<sup>18</sup> Sherry Turkle, *Life on the Screen: Identity in the Age of the Internet*. New York: Simon and Schuster, 1995, page 110.

<sup>19</sup> “*Romanticism*, an approach that dominated art during the first quarter of the nineteenth century, ... Emotional experience was prized above all, and Romanticism was both a reaction against the 'establishment' -- church, aristocratic state, and rational Enlightenment thought -- and a manifestation of the revolutionary political spirit that animated the French and American revolutions. ... The Romantic ideal of a return to nature also included *human*

## The Romantic Reaction

To more clearly explain Turkle's diagnosis and its attendant philosophical implications, I need to recapitulate an AI debate of the mid-1970s between Joseph Weizenbaum (a computer scientist and former AI researcher) and Kenneth Colby (a psychotherapist and AI designer). The Weizenbaum/Colby debate is very similar to the Lanier/Negroponte conflict and Turkle uses it in her explanation of the "romantic" position.

Weizenbaum and Colby originally collaborated on the design of a computer program that could simulate a Rogerian psychotherapist. Weizenbaum's work resulted in a computer program called ELIZA<sup>20</sup> that could carry on a textual "conversation" with someone who typed sentences and often questions to ELIZA; ELIZA would respond with various replies which were not so much answers as encouragement for the "user" to continue writing about personal problems or issues concerning family and intimate relationships. Weizenbaum became appalled by the way in which Colby and others began to see ELIZA as a first step towards automating psychotherapy. Weizenbaum is most succinct about his prescriptions for AI design at the end of his book which critiques AI:

Not all projects, by very far, that are frankly performance-oriented are dangerous or morally repugnant. ... There are, however, two kinds of computer applications that either ought not to be undertaken at all, or, if they are contemplated, should be approached with utmost caution. ... The first kind I would call simply obscene. These are ones whose very contemplation ought to give rise to feelings of disgust in every civilized person. The proposal I have mentioned, that an animal's visual system and brain be coupled to computers, is an

---

*nature.*" In Robert Atkins, *Artspoke: a guide to modern ideas, movements, and buzzwords, 1848-1944*. New York: Abbeville Press, 1993, page 185.

<sup>20</sup> Joseph Weizenbaum, "ELIZA – A Computer Program for the Study of Natural Language Communication between Man and Machine," *Communications of the Association for Computing Machinery* (9: 36-45), 1966.

example.<sup>21</sup> ... I would put all projects that propose to substitute a computer system for a human function that involves interpersonal respect, understanding, and love in the same category. I therefore reject Colby's proposal that computers be installed as psychotherapists, not on the grounds that such a project might be technically infeasible, but on the grounds that it is immoral.<sup>22</sup>

While the Lanier/Negroponete debate addresses a different AI technology than the Weizenbaum/Colby debate (the former addressing digital agents and the latter computational psychotherapists, like ELIZA), Lanier -- like Weizenbaum -- exhibits anxiety and moral outrage. Both Lanier and Weizenbaum are concerned about the ways in which machines might replace or be equated to humans and vice versa. Turkle describes this anxiety and elaborates on its manifestations in a variety of people before naming it the "romantic reaction:" "Involvement with ELIZA actually reinforced the sense that the human 'machine' was endowed with a soul or spirit -- that people, in other words, were not like computers at all."<sup>23</sup>

The "romantic reaction" does not pose a serious, philosophical challenge to AI practitioners designing computational psychotherapists and digital agents because, like the AI practitioners, Lanier and Weizenbaum assume that what is "human" is easily identifiable (at least by them) and essentially unchanged by technological "development.". Such an idealization of the human condition provides the utopic hope that "we" -- as "humans" -- could all get together and decide which technologies are "good" and which are "evil" for all humans. Unfortunately, such a utopic hope does not take into account the possibility that humans may be radically different from one another (in their needs, wants, and desires) partly because of their differing exposures and couplings to past and existing technologies.

---

<sup>21</sup> Note that examples like this, cyborgs -- the couplings of bodies to computers -- are those which have been ironically privileged by various feminist, post-structuralist theorists as an ethically sound and sophisticated picture of contemporary, human subjectivity (cf., op. cit., Haraway).

<sup>22</sup> Joseph Weizenbaum, *Computer Power and Human Reason*. San Francisco: W.H. Freeman and Company, 1976, pages 268-269.

<sup>23</sup> Op.Cit., Turkle, page 110.

Nonetheless, what is unique to romantic critiques of AI, like Weizenbaum's and Lanier's, is their engagement with ethical issues. Most of the rest of the criticism read by AI practitioners insists on what "cannot" be accomplished by AI research, rather than what should or should not be attempted.

### **Kantianism**

The "romantic reaction" of Weizenbaum, Lanier, and Turkle's subjects of the 1970s and early 1980s<sup>24</sup> was a refusal to equate humans with machines. Their refusal is a romantic one because it distinguishes humans from machines by insisting that machines do not have soul, spirit, feeling, or emotion. In other words, a romantic position identifies the *irrational* side of humans as an essential quality. As Lanier points out, materialism is one possible position in opposition to his. However, many of the opponents whom Lanier identifies are rationalists, not materialists. Chronologically, in the history of Western philosophy, the rationalists (like René Descartes) precede the romantics. The rationalists see reason and rational behavior as the quintessential human trait and so, insofar as a machine can be understood to reason more or less rationally, a rationalistic stance is not at odds with the idea of viewing a machine as an anthropomorphic entity.

It is useful to distinguish two types of rationalism in AI. The first is a crude Cartesianism in which humans are considered to be human because they can think logically or rationally. Although AI critics are fond of characterizing some AI as a Cartesian enterprise<sup>25</sup>, finding an AI practitioner who willingly responds to the label of "Cartesian", is difficult. A modified rationalism - of a Kantian, empiricist-influenced sort -- is easier to identify in contemporary AI work. This second sort of rationalism is one in which reason is considered to be central, but not sufficient to human existence; or, an understanding of rationality as bounded by certain limits (cf., AI scientist,

---

<sup>24</sup> Ibid.

<sup>25</sup> Cf. Richard Coyne, *Designing Information Technology in the Postmodern Age*. Cambridge, MA: MIT Press, 1995, page 20.

Herbert Simon's notion of "bounded rationality"<sup>26</sup>; or, more recent discussions of similar characteristics like "calculative rationality" and "bounded optimality"<sup>27</sup>).

To characterize AI-rationalism as a "Kantian-rationalism," as I am doing here, makes it easier to see how most AI work is not necessarily diametrically opposed to a "romantic reaction." Kant was interested not only in reason, but also in -- what he termed -- the other "cognitive faculties:" understanding and imagination. In other words, while reason is central to cognition, it also has its limits and interactions with other faculties; interactions which are facilitated by what Kant calls "common sense."<sup>28</sup>

Actually, Kant is more often mentioned -- not in discussions of AI -- but in discussions of cognitive science, an enterprise that is assumed to encompass AI, cognitive psychology, neuroscience, cognitive anthropology, and various works in the philosophy of mind.<sup>29</sup> Nonetheless, it is not hard to see Kant's influence on AI software design. For example, Kant's notion of "schemas" as a means for representing knowledge has been elaborated into a variety of data structures (variously called "schemata," "scripts"<sup>30</sup>, and "frames"<sup>31</sup>).

The Kantian-influenced underpinnings of many AI projects imply a certain approach to the ethics of AI research. AI critic/proponent Margaret Boden argues that, within a Kantian philosophy of morals, it would be consistent to grant rational, AI machines the same respect accorded rational

---

<sup>26</sup> Op. Cit., Simon.

<sup>27</sup> Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*. Englewood Cliffs, NJ: Prentice Hall, 1995, page 845.

<sup>28</sup> See Gilles Deleuze, *Kant's Critical Philosophy: The Doctrine of the Faculties*. Translated by Hugh Tomlinson and Barbara Habberjam. Minneapolis: University of Minnesota Press, 1984, page 21. See also the following references for different approaches to the AI sub-discipline of "commonsense reasoning," a very Kantian endeavor: John McCarthy, *Formalizing Common Sense*. Edited by Vladimir Lifschitz. Norwood, NJ: Ablex Publishing Corporation, 1990. Jerry Hobbs and Robert Moore (editors), *Formal Theories of the Commonsense World*. Norwood, NJ: Ablex Publishing Corporation, 1985. Drew McDermott, "A Critique of Pure Reason." *In The Philosophy of Artificial Intelligence*, ed. Margaret A. Boden, New York Oxford University Press, 1990. Douglas Lenat and R. Guha, *Building large knowledge-based systems*. Reading, MA: Addison-Wesley, 1990.

<sup>29</sup> Cf., Howard Gardner, *The Mind's New Science: A History of the Cognitive Revolution*. New York: Basic Books, 1985, page 56-60; Andrew Brook, *Kant and the Mind*. Cambridge, UK: Cambridge University Press, 1994, pages 12-14; and, Jean-François Lyotard, *The Inhuman: Reflections on Time*. Translated by Geoffrey Bennington and Rachel Bowlby. Stanford, CA: Stanford University Press, 1991, page 15.

<sup>30</sup> Roger Schank and Robert Abelson, *Scripts, Plans, Goals, and Understanding*. Hillsdale, NJ: Lawrence Erlbaum, 1977.

<sup>31</sup> Marvin Minsky, "A Framework for Representing Knowledge." *In Mind Design: Philosophy, Psychology, Artificial Intelligence*, ed. John Haugeland. Cambridge, MA: MIT Press, 1981.

humans.<sup>32</sup> This seems to be exactly the sort of morals that thoroughly horrify Weizenbaum and Lanier. And yet, the AI critiques written from a Kantian perspective result in the same set of questions that “romantics,” like Weizenbaum and Lanier ask: “What would happen if machines were accepted as equal to humans?” The only difference between the romantics and the AI Kantians is that the romantics are repulsed by the “very idea”<sup>33</sup> and the Kantians are not. But, neither the romantics nor the Kantians have the conceptual tools to describe how technologies influence humans (and vice versa) even when humans and machines are not equated.

### **The Aesthetic Turn**

In his book, *Designing Information Technology in a Postmodern Age*, Richard Coyne argues that AI (and computer design, in general) is moving away from rationalistic design in a move he calls "the pragmatic turn." Included in his "pragmatic turn" could be the sort of AI and AI criticism inspired by Heideggerian phenomenology<sup>34</sup>, the pragmatism of John Dewey<sup>35</sup>, and the media theory of Marshall McLuhan. I find it rather confusing to mix these philosophies together and call the result "pragmatism" as Coyne does.<sup>36</sup> Instead, if these three philosophies are to be mixed (as they undeniably are in the design of the "postmodern age"), then it might be more helpful to note that all three are concerned with the body, the senses, and the environment. These are exactly what are essential to aesthetics:

Aesthetics is born as a discourse of the body. In its original formulation by the German philosopher Alexander Baumgarten, the term refers not in the first place to art, but, as the

---

<sup>32</sup> Margaret Boden, *Artificial Intelligence and Natural Man*, Second Edition. New York: Basic Books, Inc., 1987, page 469.

<sup>33</sup> John Haugeland, *Artificial Intelligence: The Very Idea*. Cambridge, MA: MIT Press, 1985.

<sup>34</sup> E.g., Hubert Dreyfus, *What Computers Still Can't Do*. Cambridge, MA: MIT Press, 1992; and, Philip Agre and David Chapman, “Pengi: An Implementation of a Theory of Activity.” In *Proceedings of AAAI-87* in Seattle, WA, Morgan Kaufmann, 1987, pages 268-272.

<sup>35</sup> Elliot Soloway and Amanda Pryor. “The Next Generation in Human-Computer Interaction.” *Communications of the ACM* (Special Issue on Learner-Centered Design) 39 (4 1996): 16-18.

<sup>36</sup> Richard Coyne, *Designing Information Technology in the Postmodern Age*. Cambridge, MA: MIT Press, 1995, page 17.

Greek aisthesis would suggest, to the whole region of human perception and sensation, in contrast to the more rarefied domain of conceptual thought.<sup>37</sup>

In AI discourse, the “aesthetic turn” away from a strict Kantian conceptual thought (e.g., of the kind described in Kant’s first two *Critiques*) and toward an examination of the roles of the body, the senses, and the environment in cognition can be understood as the relaxation of an extreme rationalist, idealist position adopted by early-AI researchers and cognitive scientists.<sup>38</sup> AI was a critique of social sciences based in *behaviorism*.<sup>39</sup> As a cognitive science, AI began by almost entirely ignoring “external” entities and behaviors (the research domain of behaviorists) to concentrate on cognition and “internal,” mental representations.

The concerns of new “aesthetic AI” movements like connectionism<sup>40</sup>, behavior-based AI<sup>41</sup>, and distributed AI<sup>42</sup> could also be understood as a reanimation of cybernetics. So, for instance, cyberneticians’, Walter Pitts’ and Warren McCulloch’s, interests in homomorphisms between electrical circuits and the neurophysiology of the brain can be seen as a direct precursor to “connectionism,” a “new” sort of AI.<sup>43</sup> Many of the founders of AI were former students and junior colleagues of the cyberneticians in the 1940s, 50s, and 60s, so that AI might now be re-inventing cybernetics should not come as a complete surprise. In a 1988 edited collection of critiques concerning connectionism and “neural nets” AI researcher Seymour Papert recounts<sup>44</sup> -- and disavows -- the Oedipal narrative in which he and Marvin Minsky -- one of McCulloch’s

---

<sup>37</sup> Terry Eagleton, *The Ideology of the Aesthetic*. London: Basil Blackwell, 1990, page 13.

<sup>38</sup> For example, Noam Chomsky, “A Review of B.F. Skinner’s *Verbal Behavior*.” In J.A. Fodor and J.J. Katz, eds., *The Structure of Language: Readings in the Philosophy of Language*, pages 547-78.

<sup>39</sup> Op. Cit., Gardner.

<sup>40</sup> E.g., James L. McClelland David E. Rumelhart (eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition; Volumes 1 and 2*. Cambridge, MA: MIT Press, 1986.

<sup>41</sup> E.g., Rodney Brooks, “Intelligence Without Representation.” *Artificial Intelligence* 47 (1991): 139-160.

<sup>42</sup> E.g., Les Gasser, “Social Conceptions of Knowledge and Action: Distributed Artificial Intelligence and Open Systems Semantics.” *Artificial Intelligence* 47 (1991): 107-138.

<sup>43</sup> The historian of science, Steve Heims states this historical lineage between Pitts, McCulloch and contemporary connectionism. See Steve Heims, *The Cybernetics Group*. Cambridge, MA: MIT Press, 1991, page 278-279.

<sup>44</sup> Seymour Papert, “One AI or Many?” In *The Artificial Intelligence Debate: False Starts, Real Foundations*, ed. Stephen Graubard, Cambridge, MA: MIT Press, 1988.

former, junior colleagues -- co-authored a mathematical critique of connectionism<sup>45</sup> so strong that it effectively killed research in the connectionist tradition for years so that, until the mid-1980s, AI funding went predominantly to a Kantian, “symbolic” type of research.

AI criticism (produced by critics from outside of AI, e.g., philosophers and anthropologists, but also by disaffected AI practitioners) which engendered the “aesthetic turn” largely took an essentialist stance of this sort: “AI will not succeed because humans have but computers do not, and cannot, have one or more of these:” bodies<sup>46</sup>, on-going social relationships<sup>47</sup>, neurobiological brains<sup>48</sup>; and, situated, indexical representations of the surrounding environment.<sup>49</sup> AI researchers’ response to these kinds of essentialist objections has either been to dismiss them (e.g., as Herbert Dreyfus’ Heideggerian phenomenological inspired critiques of AI were hostilely dismissed for years<sup>50</sup>) or to show how computers with features, such as the following, could be designed: bodies<sup>51</sup>, on-going social relationships<sup>52</sup>, neurobiological brains<sup>53</sup>, and, situated, indexical representations of the surrounding environment.<sup>54</sup>

Critiques and AI responses of an “aesthetic”, neo-behaviorist, neo-cybernetic kind (like those listed above) are weak in the same ways in which the “romantic reactions” of Weizenbaum and Lanier are weak. From both of these perspectives the problem of AI is a problem of mimesis, i.e., to make machines which resemble humans as closely as possible. Their only difference is over whether these simulations are “real” or not. The “weakness” of these essentialist critiques of AI

---

<sup>45</sup> Marvin Minsky and Seymour Papert, *Perceptrons: An Introduction to Computational Geometry*. Cambridge, MA: MIT Press, 1969.

<sup>46</sup> E.g., Op. Cit., Dreyfus.

<sup>47</sup> E.g., Op. Cit., Winograd and Flores.

<sup>48</sup> E.g., John Searle, “Minds, Brains, and Programs.” *In Mind Design: Philosophy, Psychology, Artificial Intelligence*, ed. John Haugeland. Cambridge, MA: MIT Press, 1981.

<sup>49</sup> Lucy Suchman, *Plans and Situated Actions: The Problem of Human/Machine Communication*. New York: Cambridge University Press, 1987.

<sup>50</sup> I do not think Dreyfus intends his critiques to be essentialist, but I do think they are received as such. See Charles Spinosa and Hubert Dreyfus, “Two Kinds of Antiessentialism and Their Consequences.” *Critical Inquiry* 22 (Summer 1996): 735-763.

<sup>51</sup> E.g., Op. Cit., Brooks.

<sup>52</sup> E.g., Op. Cit., Gasser.

<sup>53</sup> E.g., Op. Cit., McClelland and Rumelhart.

<sup>54</sup> E.g., Op. Cit., Agre and Chapman.

is that they all (1) rely on an assumption that computers can (or cannot) respectably simulate humans; and, (2) ignore the ways in which ontological boundaries between humans and machines are being contested even when humans are not equated to machines.

### **Ethics, Neo-Cybernetics and Narrative**

AI's response to the "aesthetic turn" (or "reinvention" of cybernetics) in criticism has been limited due to AI's position as a sub-discipline of computer science. Notions of the body, the senses, perception, and social discourse have been introduced into newer, aesthetic, AI research, but, as I point out above, researchers have attempted to deal with these aesthetic concerns by "incorporating" them into new sorts of computer software and hardware.<sup>55</sup> Various other fields that are not bound to the disciplinary confines of computer science have also "reinvented" cybernetics. For example, the new fields of chaos theory, non-linear dynamics, and artificial life include biologists, mathematicians, computer scientists, ecologists, physicists, and others. Within, for instance, artificial life researchers have not been limited to the invention of new software and hardware, but have also begun to examine other organic and inorganic systems as systems of "life."<sup>56</sup> Artificial life can be read as an "aesthetic" critique of AI. However, these other forms of scientific reinvention of cybernetics (like artificial life) suffer from their own limitations. In these new fields, everything is seen through the "lens" of theories of chaos and complexity, which are assumed to be powerful enough to function as "master signifiers" or meta-representations.

Nevertheless, even such limited versions of neo-cybernetics can have ethical and cultural implications beyond the rational versus romantic reactions discussed earlier. The ethical implications of a cybernetics-inspired criticism hinge around this ecological observation: everyone and everything is potentially, inextricably articulated together and thus – at least in principle –

---

<sup>55</sup> Sherry Turkle speaks of what might be called AI's "incorporations," before its "aesthetic turn," in the language of psychoanalytic object-relations theory; Sherry Turkle, *The Second Self: Computers and the Human Spirit*. New York: Simon and Schuster, Inc., 1984.

<sup>56</sup> E.g., Rodney A. Brooks and Pattie Maes (eds.), *Artificial life IV: Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living*. Cambridge, MA: MIT Press, 1994.

worthy of ethical treatment. Unlike Lanier, who would reserve ethical treatment for only “what is enough like us,” a cybernetics-inspired position eschews essentialist definitions of self and other.

Social, political, and literary criticism has also “reinvented” cybernetics. Of particular interest to these theorists has been the articulation of new visions of subjectivity and “human nature” that are not based on “essences” but, instead, show how “human nature” is in constant interaction, and thus constantly changing with the environment. Jacques Lacan’s theories of subjectivity<sup>57</sup>, Gilles Deleuze’s and Felix Guattari’s notions of schizoanalysis and “desiring machines”<sup>58</sup> and Donna Haraway’s ironic, utopic vision of cyborgs<sup>59</sup> are all critical responses to cybernetics that offer anti-essentialist vocabularies for speaking about “human nature.” Unfortunately, when these new vocabularies have been used to critique AI (e.g., Manuel De Landa employs the theoretics of Deleuze and Guattari<sup>60</sup>), AI has either ignored the critiques or understood them as, in some sense, a sort of humanities branch of neo-cybernetics, like artificial life.

Some critics have encouraged equating these new theories of subjectivity to the scientific work conducted in artificial life and elsewhere.<sup>61</sup> However, as the critic N. Katherine Hayles points out scientific systems analyses can always be supplemented by a narrative which draws together the elements of any given system in ways that cannot be articulated in systems theory: in other words, systems theory is *not* a meta-representation which encompasses narrative.<sup>62</sup> Hayles makes her point by closely analyzing the rhetoric and histories of Humberto Maturana’s work as well as the narrative supplements that he was required to make to his texts.<sup>63</sup> Maturana’s texts were originally written as scientific systems analyses but, during publication and thereafter, he was required to add

---

<sup>57</sup> See Jonathan Elmer, “Blinded Me with Science: Motifs of Observation and Temporality in Lacan and Luhmann.” *Cultural Critique* 30 (Spring 1995): 101-136.

<sup>58</sup> See Brian Massumi, “The Autonomy of Affect.” *Cultural Critique* 31 (Fall 1995): 83-109.

<sup>59</sup> See Chris Hables Gray, Steven Mentor and Heidi Figueroa-Sarriera (eds.), *The Cyborg Handbook*. New York: Routledge, 1995.

<sup>60</sup> Manuel De Landa, *War in the Age of Intelligent Machines*. New York: Zone Books, 1991.

<sup>61</sup> E.g., Op. Cit., Massumi, page 93.

<sup>62</sup> N. Katherine Hayles, “Making the Cut: The Interplay of Narrative and System, or What Systems Theory Can’t See.” *Cultural Critique* 30 (Spring 1995): 71-100.

<sup>63</sup> Humberto Maturana’s work has been used by many AI scientists to re-invigorate the ideas of cybernetics within AI and in cognitive science, in general. (See, for instance, op. cit., Winograd and Flores; see also, Francisco J.

forwards, introductions, summaries, and other narrative forms. Hayles' narrative is, in my mind, a good illustration of Donna Haraway's point that all scientific and technological artifacts are not only theories and functional objects, but also myths, whether or not their inventors intend them to be.<sup>64</sup>

Victor Margolin has pointed out that the artificial/natural boundary is continually under contention and it is there, at the boundary, that the work of design and design criticism is to be done. Rationalistic, romantic and "aesthetic" critiques of AI assume that the boundary between "human nature" and "artificial intelligence" is either unbridgeable or non-existent. In other words, such critiques assign a timeless, unchanging structure to what is better characterized as an on-going struggle to negotiate the ways in which the "artificial" flows into the "natural" and vice versa. Another set of critics<sup>65</sup> has begun to articulate a cybernetics-inspired theory of "human nature" which is also sensitive to the powers of myth and narrative. These newer critiques are consistent with an argument to offer ethical treatment not only to others who are "enough like us," but everyone and everything since we may all be interconnected in strange, unconscious and chaotic ways.<sup>66</sup>

---

Varela, Evan Thompson, Eleanor Rosch, *The embodied mind: cognitive science and human experience*. Cambridge, MA: MIT Press, 1991.)

<sup>64</sup> Donna Haraway, *Primate visions: gender, race, and nature in the world of modern science*. New York: Routledge, 1989.

<sup>65</sup> E.g., Op. Cit., Gray, Mentor, and Figueroa-Sarriera; See also, Jonathan Crary and Sanford Kwinter (eds.), *Incorporations*. New York: Zone, 1992.

<sup>66</sup> Cf., Julia Kristeva, *Strangers to Ourselves*. Translated by Leon S. Roudiez. New York: Columbia University Press, 1991.