Editorial

AI and the Mechanistic Forces of Darkness

Eric Dietrich
Program in Philosophy, Computers, and Cognitive Science
Binghamton University

Under the Superstition Mountains in central Arizona toil those who would rob humankind of its humanity. These gray, soulless monsters methodically tear away at our meaning, our subjectivity, our essence as transcendent beings. With each advance, they steal our freedom and dignity. Who are these denizens of darkness, these usurpers of all that is good and holy? None other than humanity's arch-foe: The Cognitive Scientists -- AI researchers, fallen philosophers, psychologists, and other benighted lovers of computers. Unless they are stopped, humanity -- you and I -- will soon be nothing but numbers and algorithms locked away on magnetic tape.

What are the prospects of stopping these . . . these cognitive scientists? Not good; their power is enormous. They have on their side the darkest of forces: modern, Western logocentrism. Using this source, they aim at nothing less than a complete objectifying of humankind. This objectification -- this replacing of the human spirit with a computational model of mind -- is not only the most pernicious assault we humans have ever experienced, it is arguably the most insidious. It doesn't matter whether or not the objectifying world view is correct (arguments against it, even devastating arguments, apparently have no effect on it). All that matters is that it is *useful* in some limited technological sense. Why? Because, given humankind's love of technology and our ability to re-invent ourselves, cognitive science's technological success will virtually guarantee that we will re-invent ourselves as computers. I quote G. B. Madison:

> [AI]'s real significance or worth [lies] solely in what it may contribute to the advancement of technology, to our ability to manipulate reality (including human reality), [but] it is not for all that an innocuous intellectual endeavor and is not without posing a serious danger to a properly human mode of existence. Because the human being is a self-interpreting or self-defining being and because, in addition, human understanding has natural tendency to misunderstand itself (by interpreting itself to itself in terms of the objectified by-products of its own idealizing imagination; e.g., in terms of computers or logic machines) -- because of this there is a strong possibility that, fascinated with their own technological prowess, moderns may very well attempt to understand themselves on the model of a

> computational machine and, *in so doing* , actually make themselves over into a kind of machine and fabricate for themselves a machinelike society structured solely in accordance with the dictates of calculative, instrumental rationality (Madison, 1991).

In other words, it doesn't matter whether humans really are computers or not; it doesn't matter whether we can explain human thought in terms of algorithms or not, all that matters is the love affair between humans and their computer technologies. Who can resist an enemy so appealing that you want to emulate it? The power of the mechanistic forces of darkness lies in its allure. We *want* to be machines. And so we will be . . . cold, lifeless, computers, forsaking our transcendent reality, our properly human mode of existence, and instead worshiping instrumental rationality.

This attack on cognitive science saddens me, and not just because I am a cognitive scientist, and not just because I don't like being told my life's work is pointless or evil. Professor Madison belongs to a large and varied group of philosophers, stretching from Heidegger to Searle and Dreyfus, who have attacked scientific psychology (cognitive and otherwise) for one reason or another. We cognitive scientists (and indeed scientists of all stripes) have been accused of everything from killing the human spirit to just plain stupidity. Artificial intelligence has perhaps received the brunt of the attack in recent years, but all of cognitive science has been lambasted at one time or other by those who fear the human sciences. Such attacks go with the territory. Anyone who cannot stand the heat, ought to stay out of the kitchen.

No, it's not the heat. This attack on cognitive science saddens me because it is stark testimony to two depressing human traits: our tendency to oversimplify, and our steadfast refusal since the dawn of time to see ourselves as animals and therefore as part of the natural order. I am fully aware that, also since the dawn of time, there have been many who resisted oversimplification and who not only viewed themselves and all humans as part of nature, but delighted in such a perspective. If not for all of these men and women, we would have teetered on the brink of extinction centuries ago. Still, it's hard to be optimistic. Our two traits are evidenced in abundance today, and they play off of each other in a sort of positive feedback loop which shows every sign of waxing.

This essay is about the version of this feedback loop which exists between AI and cognitive science, on the one hand, and those who fear the mechanistic forces of darkness, on the other. I call this latter group *the anti-mechanists* . It is a mixed group comprising a large contingent of postmodernist philosophers as well as philosophers such as John Searle who

consider themselves part of the Western, rationalistic tradition. (To illustrate just *how* mixed this group is, the postmodernists use the phrase "Western, rationalistic tradition" derogatively.) Actually, I won't be discussing all anti-mechanists. Their numbers are legion, and they have attacked AI and cognitive science from many different directions. I am going to focus on those anti-mechanists who, like Madison, above, attack AI because of its seeming necessary attachment to logic. I need a name for this arm of the anti-mechanist army, so I will call them *the logophobes* , and their position *logophobic* . First, I will discuss AI's tendency to oversimplify. Then I will suggest that the logophobes' violent reaction to AI and cognitive science -- their dread of the mechanistic forces of darkness -- is really a reaction to this oversimplification, a reaction fueled by their refusal to view humankind as animals who are just as much a part of the natural order as cockroaches. (This refusal to see humankind as part of the natural order is what all anti-mechanists have in common.)

Logocentrism, as I said, is the source of the mechanistic forces' power, and what the logophobes fear. Logocentrism, in its anti-mechanist usage (its post-modern usage), is the love of logic: making logic or instrumental rationality central to our world-view. In the beginning, though, was *logos* , and logocentrism should be the love of logos. "Logos," the word, is a Greek noun which, in the classical period, expressed a range of meanings covered today by such words as "word," "speech," "argument," "doctrine," "explanation," "principle," and "reason." Logos, the thing, is wisdom and reason and rationality, existing not as properties of human minds, but rather as cosmic principles in a disembodied, eternal state. The ancient Greeks (particularly the Pre-Socratics, Sophists, and Stoics) enshrined logos as the controlling principle of the universe. Even up through the Judeo-Christian tradition, logos is associated with divine wisdom. But wisdom is a far cry from logic. Somewhere along the line, "logocentrism" became pejorative. Somewhere along the line, wisdom was pushed aside, and logic took its place. And I think we all know where: Logical Positivism.

Logical positivism was an immense and incredibly powerful philosophical movement of the early decades of the twentieth century. It had its roots in the nineteenth century philosophy called positivism developed by Auguste Comte. Comte held that humankind, each individual human, and indeed every branch of human knowledge grows by going through three stages. The first is the theological stage in which all natural phenomena are seen as the direct products of supernatural agents who, more often than not, are focused on humans either for good or for ill. Second, comes the metaphysical stage in which the supernatural agents are replaced by abstract, but not necessarily physical, forces. Finally, comes the *positive* stage, in which the search for ultimate causes is abandoned, and observation and reasoning are brought to bear on the task of

discovering the laws of nature. Knowledge is reconstructed in terms of experience, and the scientific world-view reigns. Science and the scientific method, according to Comte, are the high-water mark both of human culture and of each of us as individuals.

Logical positivism took off from here. With the impressive developments in logic, mathematics, and physics in the early part of this century, the philosophers, mathematicians, economists, and physicists of the Vienna Circle together with the likes of Karl Popper and Ludwig Wittgenstein (in his *Tractatus* period, and against his wishes, apparently), set out to rid the world once and for all of everything that was unanswerable, imprecise, unverifiable, and metaphysical. (The Circle included such luminaries as Otto Neurath, Rudolf Carnap, Moritz Schlick, Herbert Feigl, and Felix Kaufmann.) In short, the logical positivists set out to rid the world of all the big philosophical questions that humans have wrestled with since the beginning. They rejected all philosophical questions as meaningless.

Under the scythe of logic, fell such questions as "How ought we treat our fellow beings?" "What, if anything, can we know?" and "What is reality?" The logical positivists were intent on destroying philosophy. They proposed to replace it with science (the scientific outlook and empiricism), mathematics, and above all, logic. You have to take your hats off to these men and women: heady with the power of Frege's then recently developed formal axiomatic theory (1879), Russell and Whitehead's *Principia Mathematica* (1910-1913), and the then recent developments in science (notably biology and physics), they set out to put all of human knowledge on a basis as firm as granitic bedrock. With their new logical apparatus, they intended to define all of human knowledge in terms of verifiable, direct observation. It was a worthy endeavor. And during their heyday their influence was enormous. The successes of science and mathematics have always made everything associated with them seem utterly true and unquestionable. In our early twentieth century, this "halo effect" was especially strong in the new anti-philosophy based on science, mathematics, and logic.

Unfortunately, it didn't work. For technical reasons, some of their foundational principles were too strong, and threatened to destroyed not only philosophy but science as well. Obviously the medicine was too strong. But more importantly, they never had any success in logically reconstructing human knowledge (or even scientific knowledge) in terms of experience. The connection between what we think and what we experience is far too complicated to be captured by something as simple as logic (and logic is *not* simple when compared to other artifacts humans have created; that it is too simple to serve as the language of thought and experience is testament to how complex thought and experience are). And, finally, try as they might, the

logical positivists couldn't make philosophy disappear. Humans rationally and legitimately wonder about "the big issues," issues such as "What is the world?" and "How can we be certain of what we know?" Ultimately, logical positivism was unmasked for what it was, a genuine, but naive attempt to simplify the world in a way in which it couldn't be simplified. Philosophy is nothing if not difficult precisely because it tackles the big questions, and the big questions are very messy. The logical positivists failed to appreciate the vast gulf between what they wanted to revamp (human knowledge) and the tools they had for revamping it. In short, logical positivism attempted to *oversimply* the world and our knowledge of it.

Logical positivism, as a philosophy, had completely faded sometime around the mid-1950s (or possibly as early as the 1940s or as late as the 1960s, historians disagree). But, as is well-known to philosophers, especially philosophers of science, logical positivism didn't really die. Like a contagious disease, it moved out of the philosophy departments and down the halls to our less immune colleagues. In the transfer, it became attenuated; it lost its high-minded goal of making human knowledge solid and secure, while retaining the goal of couching everything in logic so that all could be rendered perspicuous and easily dealt with.

Science has gone on to embrace the complexity of nature. Chaos theory and the rising sciences and mathematics of complexity are a bright, invigorating testament to this. And nowhere is embracing nature in all its complexity more important than in psychology and cognitive science, for the brain is arguably the most complicated device on the planet. And yet nowhere is the existence of logical positivism more obvious. Cognitive science, especially artificial intelligence, is riddled with it. The sad truth is that logical positivism, in its attenuated form, has deeply infected artificial intelligence.

I offer as evidence for this claim the quantity of research on logic that goes on in philosophy, and the percentage of all AI research that focuses on logic. Fully one-half of the papers Jetai receives report some logical result or other. There are entire conferences, books, and journals devoted to logic in AI. Much of this research is fascinating, important, and useful for unraveling and furthering our understanding of the vast domain that is logic. AI researchers have invented whole new kinds of logics, logics with very strange axioms whose behavior is quite unexpected. But at the end of the day, even with all our nonmonotonic, quantified, modal logics, we have not succeeded in explaining one iota of human cognition; at best we've merely described it. We have said *what* intelligence is; we haven't said *how* it occurs. And I say this knowing full well that AI logic researchers believe that they have almost completely unraveled the mystery of one aspect of human cognition: commonsense reasoning -- the sort of reasoning we all use

everyday to get through the world. But, alas, they have not. For starters, the things that get along best in the world are not robotic implementations of the nonmonotonic logics describing common sense, but the "situated" robots such as those being built by Rodney Brooks at MIT. (I refer the reader to Lynn Stein's 1994 paper in Jetai, vol. 6.4.)

Secondly, it is not even clear that commonsense reasoning is a natural cognitive kind and that it should be getting the attention that it does. Commonsense reasoning considered just as a phenomenon strikes me as a surface feature of much more complicated, lower-level computations. Every time I read an article on common sense reasoning, I am reminded of a young geology student I knew who rocketed out of camp in the Snowy Range early one morning, mapping the boulders he saw, completely forgetting that boulders "float" on the surface soil. They have tumbled from nearby mountains, and are pushed along by erosion. Boulders are not indicative of the bedrock underlying the earth. To get at that you must carefully map outcroppings (not floating boulders) and fold in seismic and borehole data. I think common sense researchers are mapping "float." Consider for example, the greatest power of human cognition: creativity. Common sense reasoning has nothing to say about this. Researchers of common sense seem to assume that creativity is a minor aspect of human cognition and that they can somehow unravel common reasoning without saying anything about creativity. It is much more likely that a theory of human and machine cognition worth its salt will give a central place to creativity and explain it and "common sense" as the joint products of a deep, underlying architecture.

The twin beliefs that commonsense reasoning is real and all but explained are a symptom of the hold logic has on AI. There are many others. I will cite one more. AI logic researchers know of an important case where logic failed to explain some very human behavior, yet they persist in still using logic. I am speaking of the Davidson project. In the nineteen sixties and seventies, most philosophers of language labored at what was called the Davidson project. The goal of this project was to supply English with a semantic theory. Just as Chomsky, his colleagues, and their students were to supply a recursive syntactic theory for English (and all other natural languages), so were the Davidsonians to supply a recursive theory which specified for any sentence in English, what that sentence meant. The theory, according to Davidson, was to take the form of a truth definition. That is, for any sentence S in English, the semantic theory would specify the conditions under which S was true. Where did Davidson get this idea -- the idea that meanings were truth conditions? From logic. In logic, meanings *are* truth conditions. So the Davidson project was founded on the hypothesis that English really is a disguised logical language. Philosophers of language engaged in the Davidson project, therefore, spent their time trying to

discover which logic English really was.

The Davidson project was actually a great moment in philosophy of language, in particular, and in philosophy, in general. Philosophers began with a prima facie plausible hypothesis -- natural languages were (or could be fully described as) logic. They then set about to prove this. It is now widely known -- if not widely acknowledged -- that the Davidson project gloriously and spectacularly failed. The obvious conclusion is that English is not a disguised logic, and neither are any of the other natural languages on planet Earth.

And now a modest suggestion to my fellow AI researchers: Isn't the failure of the Davidson project compelling evidence that logic's role in human cognition minimal at best? Isn't it now high time we consider the proposition that logic is all but useless in our quest to develop a theory of human and machine cognition? What to replace logic with is a very contentious question and one, thankfully, we must discuss at another time.

Now to the other half of the feedback loop. AI researchers oversimplify when they use logics of one form or another to describe cognition. The logophobes react to this, first and foremost. Logophobes fear turning humans into logic machines, as Madison's quote, above, makes clear. The force of darkness they see is logic -- logocentrism. Their fear is justified, it seems to them, for two reasons. First, as I have discussed, AI researchers do in fact spend a lot of energy devising logical descriptions of human cognition, and thus focus on those aspects of human cognition, like commonsense reasoning, where this logic agenda might in some sense succeed, concomitantly ignoring those very important aspects, like creativity, where logic is obviously of little use. Second, logophobes know that deep down in the guts of every computer, exists some Boolean logic governing its behavior. They think this fact is more important than it actually is. I will turn to this second reason momentarily.

The logophobes' first reason would make some sense if AI was completely based on logic, but it is not. AI is a lot more than just logic. If half of Jetai's submissions are logic-based, the other half are not. We get papers on genetic algorithms, connectionism, systems, distributed AI, heuristic search, situated cognition, artificial life, robots, and the ever-popular, scruffy, knowledge-based, cognitive modeling. Why would logophobes attack AI (and cognitive science) for its commitment to logic if not all of AI is committed to logic? I'm not sure. Maybe their hatred of logic blinds them to the other parts of AI. On the other hand, maybe they are so opposed to viewing humans as fancy cockroaches that they intentionally lump all of AI under the logocentric

umbrella because they know it is easier to attack that way. There is some evidence that logophobes are using this latter strategy. The non-logic AI literature is quite large. It is pretty hard to miss, accidentally. It is much more likely that the logophobes are purposely ignoring this literature.

Now for the logophobes second reason for fearing logocentrism: the idea that computers are logic machines. To begin with, computers are *not* just logic machines, and they are *not* just number crunchers. This is an extremely deep point, and one that people outside of the professional computer culture routinely misunderstand. Though I cannot do the details justice here, I will say a few things. A computer, any computer, comprises a hierarchy of *virtual machines*. Each machine is different from the one below and above it, and each supervenes on the one below it. And most importantly, each virtual machine has a methodology and mode of explanation unique to it (for example, the explanation and debugging of your word-processor is very different from the explanation and debugging of your operating system, or your disk drive). These methodologies and modes of explanation *cannot* be reduced those of the machines below. To say that everything reduces to Boolean logic in a computer is exactly like saying psychology and biology reduce to physics. The claim is true in a technical, abstract sense, but it is epistemologically wrong. To try to reduce the behavior of, say, a connectionist system to the Boolean algebra of the gates in the supporting chip would be to completely lose what was important about the connectionist system in the first place. To press the comparison, the science of life is biology, not physics. If you try to reduce biology purely to physics you are going to wind up with an incomprehensibly complex mess, and you will be methodologically stymied.

Now, it is true that computers cannot do everything -- in fact, there are important mathematical proofs that they cannot do everything. The reason AI is a deep scientific enterprise -- what makes it not vacuous -- is that everyone in AI (and cognitive science, for that matter) is committed to the claim that what computers can do *subsumes the functions that explain and implement cognition*. (I have elided an enormous amount of detail here. I suggest reading Chalmers' "On implementing a computation." And though it is a bit self-serving, I also suggest reading my book *Thinking Computers and Virtual Persons*.)

So, logophobes have ignored, perhaps willfully, the vast part of AI that is not based on logic, and they have completely misunderstood the nature of computers and computation. They have vented their ire at all of us in cognitive science, when they should be focusing only on some of us -- the logic researchers. Finally, they are correct about one thing: logic *is* inadequate for explaining cognition. It is interesting to note, that according to my analysis here, the AI logic

researchers and the logophobes share something in common -- they both misunderstand the role and importance of logic: they both think it is more important than it is.

I don't think my pointing out their error will make the logophobes feel more warmly towards cognitive science. Ultimately, I believe they are deeply opposed to viewing humans as anything "less" than spiritual, transcendent beings. They are, in short, deeply opposed to a science of humankind, and a scientific explanation of our most compelling attribute: our minds. As we finish the twentieth century, and move into the twenty-first, the false dichotomy of "mere machine" or "divine creation" continues to haunt us, bedeviling our efforts to understand ourselves as we really are -- glorious machines whose complexities make us precious.*

* I thank Clay Morrison, Larry Roberts, and Alan Strudler for comments on previous drafts of this essay.

References

Chalmers, D. (1994). On implementing a computation, *Minds and Machines* , December, 1994.

Dietrich, E., ed. (1994) *Thinking Computers and Virtual Persons* . Academic Press: San Diego.

Madison, G. (1991) Merleau-Ponty's Deconstruction of Logocentrism, in M. Dillon ed. *Merleau-Ponty Vivant* . SUNY Press: Albany. [In the table of contents, this paper is referred to as "Merleau-Ponty's Destruction of Logocentrism."]

Stein, L. (1994) Imagination and situated cognition, *J. of Experimental and Theoretical AI* , 6.4, 393-407.