Editorial

AI and the Tyranny of Galen, or Why evolutionary psychology and cognitive ethology are important to artificial intelligence.

Eric Dietrich, Dept. of Philosophy, Binghamton Univ.

Concern over the nature of AI is, for the tastes many AI scientists, probably overdone. In this they are like all other scientists. Working scientists worry about experiments, data, and theories, not foundational issues such as what their work is really about or whether their discipline is methodologically healthy. However, most scientists aren't in a field that is approximately fifty years old. Even relatively new fields such as nonlinear dynamics or branches of biochemistry are in fact advances in older established sciences and are therefore much more settled. Of course, by stretching things, AI can be said to have a history reaching back to Charles Babbage, and possibly back beyond that to Leibnitz. However, all of that is best viewed as prelude. AI's history is punctuated with the invention of the computer (and, if one wants to stretch our history back to the 1930s, the development of the notion of computation by Turing, Church, and others). Hence, AI really began (or began in earnest) sometime in the late 1940s or early 1950s (some mark the conference at Dartmouth in the summer of 1957 as the moment of our birth). And since those years we simply have not had time to settle into a routine science attacking reasonably well understood questions (for example, many of the questions some of us regard as supreme are regarded by others as inconsequential or mere excursions).

Moreover, we cannot simply assume that we will gradually settle into the mold of a standard working science for the simple reason that there are important and serious doubts that a science of intelligence and intelligent behavior in general is in the offing. (By "general" I mean that most AI researchers believe that AI theories will apply to any intelligent information processor, whether silicon-based or not.) I am not referring to the well-known philosophical doubts caused by concerns about what is or is not computationally possible -- the sorts of concerns raised by John Searle in his Chinese Room Argument, for example. Rather I mean the deep methodological concerns raised by connectionism, evolutionary computing, situated action, and artificial life (of course, some of the concerns raised by these latter approaches to machine intelligence are related to philosophical problems raised by the likes of Searle and others). It might turn out, for example, that there is nothing to say about intelligence independently of what can be said only about humans and human brains. In that case, there would be no AI distinct from neuro, cognitive, and developmental human psychology. Or it may by that the only thing that can be said about intelligence is what can be said about intelligence in individual species: dolphin intelligence is different from chimpanzee intelligence which is different from arachnid intelligence, etc. etc. (Another version of this scenario is that the various kinds of intelligences within and between species might not cohere enough to fall within the purview of one scientific theory. In this case AI might fail in the way that general systems theory failed: there just isn't much one can say about all systems in general, because systems differ from each other so much and these differences matter a great deal.) A third possibility is that building an artificial intelligence distinct from building an artificial life form is simply not possible. In this case, AI would be swallowed up by biology, both natural and computational. A very real possibility here is that we will never build an intelligent machine, rather we will build various evolutionary machines which in turn will evolve machine intelligence for us. Then in order to figure out what has been built and how it works and thinks, we will all have to become cyber-biologists -- dissecting and theorizing about computational fauna in the same way that our colleagues in lab coats theorize about guppies and bumble bees, birds and squid.

I suspect many readers are skeptical that these three are real possible futures for AI. As one antidote for this (complacent) skepticism, I suggest reading *Computational Intelligence: Imitating life* by Zurada, Marks, and Robinson.

In sum, AI is young; its future uncertain, so raising concerns about the nature of AI is entirely appropriate (in fact, we should do it routinely). In this editorial I would like to discuss some foundational concerns that have troubled me for the last couple of years. I am not going to argue that a certain methodology is unfruitful or that answering certain questions is taking us off course; these sorts of criticisms are complicated and deserve to be discussed at full length on their own. Rather I want to suggest that AI take seriously results from two fields seemingly irrelevant to AI: cognitive ethology and evolutionary psychology. (Cognitive ethology is the study of the minds of other animals; evolutionary psychology is the study of the phylogeny of minds -- mainly human minds.)

(Throughout the rest of this discussion, by "AI" I shall mean classical, symbolic, knowledge-level AI together with its emphasis on classic methods such as logic, knowledge representation, and search. When I mean AI in a broader sense -- classical AI together with neural nets, evolutionary computation, artificial life (what is sometimes called “computational intelligence”) -- I shall explicitly say so.)

Where do AI scientists get the algorithms they implement? A variety of places. Perhaps most frequently, they are generated by thinking about how to solve a certain problem. Frequently these problems are presented by some real world domain (e.g., air traffic control). Sometimes problems or algorithms are suggested by working with already existing systems (e.g., after working with an AI system for a while you can just see what other capabilities have to be added to improve the system.) Algorithms are sometimes derived from cognitive psychology (e.g., research on analogical thinking). (When this strategy works, it produces important programs which enhance both AI and cognitive psychology. The problem is that the strategy doesn’t often work. For a variety of reasons, the flow of ideas from psychology to AI is little more than a trickle. One reason is that cognitive psychologists are almost as much in the dark about minds and intelligence as we are.) Other times, AI researchers derive algorithms by varying other algorithms that are already known to work or to have interesting properties.

But all of this misses the real answer. For the most part, AI researchers dream up their algorithms, basing them on introspection, intuition, or analogies with formal systems (e.g., Post production systems or first-order logic). (Remember, I am talking of classical AI.)
Once while discussing connectionism, a classical AI colleague said to me “It just seems to me that thinking must be operations on mental symbols. That’s what I implement.” It struck me that my colleague’s work was based primarily on his intuition about what thinking must be. I think this reliance on intuition must be widespread. Of course, refinements to the algorithms come from experiments and field testing, but in the beginning, algorithms are frequently created out of whole cloth. (I am not disparaging this method; it is fine as far as it goes. But it apparently doesn’t go far enough for many researchers. This explains, partly, the appeal of deriving algorithms from other sources such as the behavior of organisms (artificial life), neurons (artificial neural nets), and evolution (evolutionary computing). I will suggest below that none us should regard the whole cloth method as going far enough.)

But now notice how odd AI’s research strategy is. We are all -- classical and nonclassical AI researchers alike -- in the business of building intelligent machines and theorizing about the phenomenon of intelligence. Yet we ignore almost all of the intelligence in the world by ignoring all of the other animals on the planet as well as our very intelligent ancestors: Homo erectus and Homo habilis. Instead we derive algorithms from humans (via introspection and from psychology and neuroscience, with limited success) and from global processes such as evolution and population flux. We leap from the most intelligent systems on the planet (us) to systems for whom "intelligent" isn't even an appropriate adjective, skipping everything in between. This failing to consider relevant data lying right under our noses reminds me of an era in the history of medicine: what has been called the tyranny of Galen (see Boorstin, chapter 45, pp. 344-350).

Galen was an ancient Greek physician (c. 130 - 200) who wrote several important books on medicine, the most influential of which was called *On the Usefulness of the Parts of the Body* . Galen's influence was due partly to the fact that he was one of the first experimental physicians, and he constantly urged his colleagues to learn from experience and to focus on knowledge that was useful, that could cure patients. But Galen's influence went beyond anything he could have imagined, and, in one of the cruel ironies of history, beyond anything he would have wanted. For approximately fifteen hundred years, until the late seventeenth century Galen's books were regarded as sacred texts. Instead of furthering the field of anatomy by dissecting human bodies, physicians read Galen. For all those centuries, medicine was not a science, but a branch of philology. Anatomy and diagnosis were done by studying what Galen said, and when what he said didn't fit the facts, physicians redoubled their efforts to figure out what he really meant. Physicians ignored the data that were right in front of them: bodies.

I think our situation in AI is analogous to this "tyranny of Galen." We aren't focusing on sacred texts, of course, but we are ignoring important data we have ready access to. We AI researchers are like the physicians of the Middle Ages, and their appeals only to the works of Galen are our appeals to everything but what is probably the most relevant data to be had, namely the phylogeny of our own minds and the minds of other animals. (One important disanalogy here is that Galenite physicians ignored real humans, whereas I am suggesting that AI researchers focus on them too much.)

It might be interesting to speculate on why we ignore the minds of other animals and the minds of our extinct ancestors. A detailed analysis will have to wait for another day, but certainly the natural isolation caused by disciplinary boundaries explains some of why we don't talk to cognitive ethologists and evolutionary psychologists. That's one reason for this essay. Also, I want to stress that in equating the situation in AI with the situation in medicine from the Middle Ages I do not mean that we should abandon any of the strategies in classical AI or in nonclassical AI. I said above that I was not going to criticize any particular methodology nor any particular set of questions or research areas. I am not urging that AI scientists drop what they are doing. Rather I am urging that we keep doing what we are doing, but with an eye on the intelligence in other animals and the evolution of intelligence in humans.

I now want to turn to two examples of how the broader perspective urged here might help us. One example will illustrate why we should study cognitive ethology and the other example will illustrate why we should study evolutionary psychology.

The primary reason to study the minds of other animals is that it would allow us to compare and contrast; that is, it would allow us to ask "in what ways are we smarter (and less smart) than other animals, and why?" Other animals simply provide us with more examples of intelligence, but since these will differ from our own, we should be able to isolate the differences and then explain their architectural causes.

Consider the East African vervet monkey (*Cercopithecus aethiops*). Vervet monkeys exhibit a striking lacuna in their knowledge of the world: though they adeptly manipulate each other and their social situations, they are incapable of rudimentary inferences regarding their predators. Vervets can feign respect and affection to further their own ends and those of their family, they retaliate for past wrongs -- sometimes picking on a relative of the wrong-doer, and they seem to understand other monkeys' status and kin relationships that do not involve them directly. Yet they cannot figure out that a winding path of recently matted down grasses is the telltale mark of a python. They will follow such a path until they walk right up on the snake, who spares them only if it's not hungry (for more details see *How Monkeys See the World* , Cheney and Seyfarth, 1990).

Of course, vervets can recognize their predators; they don't mistake a python for a tree branch. What they seem incapable of doing is recognizing typical *evidence* that a predator is near-by. Such evidence is called "secondary predator cues." In the language of primatologists, vervets seem unable to recognize secondary predator cues. Matted grasses are a secondary predator cue suggesting that a python has recently been through the area. Other predators leave other kinds of clues. Vervets are ignorant of them all.

Vervets have a kind of "laser-beam intelligence" or "tunnel intelligence." Within the social domain they are adept problem solvers, but outside of that domain they are not smart. This tunnel intelligence is genuinely puzzling. Vervets, like most primates in the wild have a tenuous hold on life, and predators are the primary reason. Predation accounts for 70% of all mortality among vervets (Cheney and Seyfarth, 1990). The numbers of vervets that become meals, suggests that vervet ignorance is architectural. They do not simply lack the requisite knowledge in the sense that they could learn this information in the right setting. They *are* in the right setting. Rather, they are probably *incapable* of acquiring the knowledge, for architectural reasons.

This case is intrinsically interesting. Vervets are capable of associative learning, yet they can't seem to associate predators with secondary predator cues. Why? There are several possible explanations of vervet brittleness, and no agreement as to the right explanation (see the precis of Cheney and Seyfarth's book in *Behavioral and Brain Sciences*, vol. 15, n. 1,1992). Importantly, some of the explanations are directly relevant to AI. First, some of them are testable using AI programs, and (and this is my main point) some of them could inspire new classes of algorithms for implementing intelligence. As one example, some ethologists have speculated that vervets' knowledge of their world is divided up into separate knowledge bases and that these bases are *inaccessible* to each other in certain ways. The algorithms vervets have which enable them to manipulate their social environment by making clever inferences are simply not available to the knowledge base storing information about predators. On the other hand, human brains also appear to store knowledge of environments in separate knowledge bases, but our knowledge bases are *accessible* to one another. Chimpanzees seem to lie somewhere between vervets and us. Perhaps, therefore, and contra received dogma in AI, humans are smart not so much because we store large quantities of knowledge, but because our knowledge bases have "open borders" allowing inference procedures available to one to be available to all. This isn't the place to pursue the details, so I urge the reader to read the precis of *How Monkeys See the World* mentioned above.

My second example illustrates why we should study evolutionary psychology: the study of our brains and minds from an evolutionary perspective. Evolutionary psychologists begin with the fact that the brains we are using now evolved in the Pleistocene, approximately 2.5 million years ago. Their most important methodological assumption is that brains (like every other animal organ) are adaptations evolved to solve various problems, and that this fact is tremendously relevant to understanding thinking and intelligence. More specifically, evolutionary psychologists use theories of selection pressures to generate hypotheses about the architecture of the human mind (and sometimes the minds of the other higher primates). And this is, in fact, the primary reason to study evolutionary psychology: it would allow us to generate richer, more robust architectures for intelligent machines, and it would provide constraints on our theories of intelligence. Since many evolutionary psychologists are good computationalists, they and we ought to be able to carry on a dialogue. Here now is my second example.

In his paper, “The phylogeny of rationality,” John Pollock argues that the architecture of cognition (and specifically the architecture of rationality -- what distinguishes us from what Pollock terms “moderately sophisticated animals”) can be derived from first principles; that is from just getting clear on how rationality ought to work given certain beginning definitions. (So Pollock’s use of the term “phylogeny” in his paper is idiosyncratic, to say the least, since the evolutionary history of cognition and rationality is not discussed.)

Pollock begins by noting that humans have many built-in, fast, task-specific procedures for figuring out our environment and acting on it accordingly. In keeping with current terminology, he calls these procedures “modules.” Modules are crucial for animal survival because rational thought is too slow for many tasks. He cites a “trajectory module” as an example (p. 566). We don’t rationally calculate where a ball is going to land, we can just see where it will go. If we had to rationally think about it, we would still be thinking long after the ball had dropped to the ground. There are many other examples. Modules are not part of our rational architecture because they gain their speed by being inflexible -- hardwired, if you will. That is why they are not part of our rational architecture -- rationality is not inflexible. Pollock calls these modules “quick and inflexible modules” or “Q&I modules” for short.

Pollock conceives of a rational agent “... as a bundle of Q&I modules with ratiocination [i.e., rational thought] sitting on top and tweaking the output [of the modules] as necessary (p. 566).” Pollock’s view of our cognitive architecture accords quite well, I think, with the received view of cognition in AI as well as in a sizable corner of cognitive science. But it might not be correct. Recent research in evolutionary psychology is quite at odds with Pollock’s conception of our minds.

Briefly, evolutionary psychologists theorize that human minds are a vast collection of modules (they are more or less self-contained, but it is not clear how inflexible all of them are). These include modules for mate selection, habitat choice, behavior towards relatives and non-relatives, and how to avoid being cheated in social exchanges, as well as the standard modules for capacities like vision and trajectory calculation. (I urge the reader to check out *The Adapted Mind* by Barkow, Cosmides, and Tooby.) The hypothesis is (and there is intriguing evidence for this) that there are *no* domain-general cognitive mechanisms satisfying the philosophers’ notion of rationality, and that the “limited generality” we are capable of is probably due to interacting modules. There is, in short, no separate mechanism for rationality of the kind postulated by Pollock which sits on top of a bunch of non-rational modules adjudicating their output.

Rationality is an emergent property of human beings (and other animals), not a property for which there is a specific hardware mechanism.

I can't resist pointing out another result from evolutionary psychology. The evidence apparently indicates that when it comes to probability reasoning, humans are frequentists, not Bayesians. If this is right, then AI projects founded on the Bayesian approach to probability will result in architectures completely different from human minds. This might be a good thing, I don't know. But it is interesting given that AI researchers are always pointing out that they don't care whether their intelligent machines mimic human architecture or not. Still, trying to mimic human architecture might not be such a bad thing; whatever it is, we know it works.

These two examples show, I think, that there is a vast world of research going on which is investigating intelligence, the mind, and cognition, a world which we are ignoring at our peril. I think that if AI (both classical and nonclassical) continues to ignore this research it will simply become irrelevant and slip slowly beneath sea of human inquiry, coming to rest on the bottom with other failed enterprises such as Ptolemaic astronomy and the caloric theory of heat.

I am urging that we broaden our horizons, and that we renew our attack on the problems of machine intelligence and cognition with open minds, breaking off frequently to discuss our research with ethologists and evolutionary psychologists and anyone else who cares to talk to us.

I close with one of my favorite quotes about open-mindedness in cognitive science from one of my favorite philosophers, Jerry Fodor: "I wish I could learn to be less discriminating still, for I am morally certain that real progress will be made only by researchers with access to an armamentarium of argument styles that considerably transcends what any of the traditional disciplines offer."

References

Barkow, J., Cosmides, and L., Tooby, J. (1992). *The Adapted Mind* . New York: Oxford University Press.

Boorstin, D. (1983). *The Discoverers* . New York: Vintage Books.

Cheney, D. and Seyfarth, R. (1990). *How Monkeys See the World*. Chicago: University of Chicago Press.

Fodor, J. (1981). *Representations*. Cambridge, MA: MIT Press.

Pollock, J. (1993). “The Phylogeny of Rationality” *Cognitive Science* **17**, 4, pp. 563-588.

Zurada, J., Marks, R., and Robinson, C. (1994). *Computational Intelligence: Imitating life*. New York: IEEE Press.