

Thinking Architectures: Symbolic vs. Connectionist AI

Rowin Andruscavage

February 27, 2001

S&TS 438 SP01 Paper 1

Can machines think? The rapid development of the computer as a general purpose information processing device has led people to wonder how long they will remain the most intelligent beings on Earth. After all, computers have already begun to beat human experts at thinking-oriented games such as chess. Will mankind soon become subordinate to a superior artificial intelligence? Not so soon, say critics of current AI techniques.

Two kinds of AI architectures now exist: the classical symbol manipulating machine, and the connectionist neural network device. The former represents the current paradigm of computing - state machines capable of satisfying the Turing test by representing any combination of possible states. Philosophers largely agree that such machines will never think consciously as humans do - no matter how human like they appear to speak and behave, their thought processes consist of nothing more than "explicit rules" applied to "atomic facts" and completely devoid of semantic understanding. Hubert Dreyfus additionally outlines the inadequacies of classical computers in achieving artificial intelligence, implying the need for more advanced architectures. Following his work, Paul and Patricia Churchland, suggest that the potential thinking capability of a connectionist neural network may far

exceed that attainable by classical computers.

The limits of classical computer AI have not always been evident. Alan Turing made popular the concept that any sufficiently flexible machine could solve any calculable problem given enough time and resources. The notion that a computer could take in any combination of stimuli and calculate the same set of responses that a human would give led to the advent of the Turing test of intelligence. Since the notion of "thinking" has always been very much a subject of debate, they had no choice but to take the behaviorist view of setting human speech and behaviour as a standard to which we compare all forms of thought. Thus, we would rate a computer's "intelligence" by how indistinguishable its responses to stimuli were from those of a human. Of course, such systems tended to work on the hypothetical principle that all possible human responses to all possible stimuli could be stored or computed in advance - not just a physical impracticality, but also a far cry from what probably occurs in the human brain. A machine that "thinks" in this fashion is the same kind of machine that predicts the stock market by looking at nothing but streams of stock values. Saying that a machine exhibits consciousness in this manner is as absurd as saying that one's reflection in the mirror has consciousness, even considering that the mirror might have sophisticated record, playback, and body part compositing capabilities. The hopes and dreams presented by so-called "Strong AI" technology promise to never become more than just an encyclopedic toy.

Many philosophers, including Searle and Collins, dismiss parallel connectionist AI as simply a collection of classical computing devices, and thus would hold to the same philosophical limits on consciousness and understanding. The Churchlands, in their neurophysiological reverse engineering studies of the brain, have gained an appreciation for this "most complicated and sophisticated thing on the planet." Of course, the fact the the human brain itself appears to consist of a large neural network provides the most compelling reason to believe a similarly fashioned connectionist AI could achieve

intelligence. Next is the sheer size and complexity : "the human brain has 10^{11} neurons, each of which averages over 10^3 connections." Furthermore, neurons deal with input and output signals unlike any used by analog or digital electronic circuits. Neurons fire discrete pulses through their axons at a frequency proportional to the weights of the inputs coming in at their various dendrites. This might allow the neuron to not only respond to differences in the frequencies of two incoming signals, but also the phase. Neural nets also tend to have feedback loop arrangements, which "allow the brain to modulate the character of its sensory processing" and completes the "genuine dynamical system" of the brain. The continuum formed by these circular groups of connections empower the brain to engage in highly complex operations that are "to some degree independent of [their] peripheral stimuli." Neural networks grow and learn by physically changing the weights and numbers of their connections.¹ Due to the difficulty in duplicating this growth in hardware, implementations of connectionist AI have generally lagged behind its symbolic counterparts, which thrive on relatively inexpensive commodity computer systems.

Connectionist AI bears the solution to many of the limitations of classical computing. Dreyfus makes this apparent in his comparison of expert systems with their human counterparts. Expert system software requires explicit programming of the rules and heuristics used to compute the solutions to their problems. However, Dreyfus notes that in human expertise, strict rules are only used by beginners. Proficient humans somehow internalize their decision making, their "brain does not work like a heuristically programmed digital computer applying rules to bits of information." Instead, human thought seems to work "holographically, superimposing the records of whole situations and measuring their similarity." Neural networks intrinsically have the ability to perform this type of template matching, where higher similarities would stimulate more neuron firings.²

¹Churchland & Churchland, "COULD A MACHINE THINK?" pg. 36

²Dreyfus & Dreyfus, FROM SOCRATES TO EXPERT SYSTEMS: THE LIMITS OF CAL-

Indeed, Dreyfus recognizes the inadequacies of an SM system's "recognition of ordinary spatial or temporal objects" as "checking off a list of isolable, neutral, specific characteristics." The list falls into the age-old problem of sorting out the relevant facts from the background of information. Parallel networks operate on the Gestalt notion of the whole being more than the sum of its parts: a sequence of tones are "perceived as part of [a] melody" rather than "independently identified notes". As the brain absorbs the notes, the part of the neural net that stores the melody reaches a state of excitation higher than those which store other melodies, perhaps stimulating some of the emotions and memories associated with the whole of the song.³ Furthermore, Dreyfus identifies various bodily skills which provide a complicated computational challenge for serial processors but could easily be abstracted by a trainable neural net. The optimal movement of highly articulate mechanical arms and manipulators "have so far proved an insurmountable obstacle" for programmers.⁴ However, a neural net with the appropriate "configuration of synaptic weights" could effortlessly compute the second order differential equations required for applying the right signals to the right actuators to move an appendage without hitting any obstacles in its path.⁵ To Dreyfus, embodiment provides the essential role of providing an experience to complement our consciousness. The extended nervous system acts as a flexible abstraction layer between our physical and mental existence.

Searle may attempt to liken a connectionist network to "an elaborate set of water pipes with valves connecting them," with the valve synapses operated by a programmed agent, ostensibly to show that a classical non-understanding Turing machine could operate such a parallel network.⁶ Likewise, Collins sees a "continuity between neural nets and the rest of AI."

CULATIVE RATIONALITY pg. 341

³Dreyfus, WHAT COMPUTERS STILL CAN'T DO: A CRITIQUE OF ARTIFICIAL REASON pg. 238

⁴*Ibid* pg. 251

⁵Churchland & Churchland, "COULD A MACHINE THINK?" pg. 36

⁶Searle, MINDS, BRAINS, AND PROGRAMS pg. 421

They're just a "super programming language" that automatically learns stimulus-response directly from the environment rather than from the programmer.⁷ Both philosophers see complex parallel systems as functionally equivalent to a classical serial processor iterating over the various components. This may be true for a synchronous neural network, where all the components operate in time with a metronome signal. The system always has a definite, consistent state after allowing an appropriate interval to complete synaptic action. However, the human brain has no such clock, and functions asynchronously. Every neuron fires whenever it needs to, so at any given moment the brain's state could include any number of neurons just about to fire, in the action of firing, or just having fired. Due to the chaotic complexity of the brain, this uncertainty vastly expands the number of state-spaces the brain is capable of, meaning the neural network has indefinite operating parameters that exceed the limits of the state-constrained Turing machines. These are not classical computers as we know them.

Collins accuses Dreyfus of embracing neural nets simply because they share a mutual enemy in 'Good Old Fashioned AI'.⁸ Dreyfus demonstrates an acute awareness of the distinction between the two. He finds it ironic that computers are the epitome of 2000 years of platonic reductionism, excelling in performing the "so-called higher rational functions - those which were once supposed to be uniquely human." Yet, "it is the sort of intelligence which we share with animals, such as pattern recognition" and communication, that has eluded the grasp of AI.⁹ Humans have long driven themselves to discipline their sloppy connectionist minds to perform symbolic logical relations. Computer intelligence has started from the human ideal and must now bend its circuits to embrace the animal condition.

⁷Collins, "EMBEDDED OR EMBODIED?" ARTIFICIAL INTELLIGENCE #80 1996 pg. 14

⁸*Ibid* pg. 13

⁹Dreyfus, WHAT COMPUTERS STILL CAN'T DO pg. 237