Artificial Intelligence and Cognitive Systems

Pat Langley

Computational Learning Laboratory
Center for the Study of Language and Information
Stanford University, Stanford, CA 94305

I became involved in AI during the 1970s, when I was in graduate school, because I wanted to understand the nature of the mind. This seemed like one of the core questions of science, on an equal footing with the nature of the universe and the nature of life. Artificial intelligence, with its computational metaphor, offered the only clear course for tackling this challenging problem, and the progress made in the field's first 20 years seemed impressive enough to promise rapid progress toward a computational theory of mental phenomena.

When I arrived at Carnegie Mellon University in 1975, and for the next 15 years, AI research drew upon a number of assumptions about the field's goals and the approaches that might achieve them. In this essay I review these assumptions, the reasons they made sense, and the additional reasons, many sociological, why they have fallen into disfavor among many AI practitioners. After this, I consider whether they have a role to play in the next 50 years of the field and, if so, how we can encourage their increased use. I will refer collectively to these assumptions as the paradigm of *cognitive systems*.

One key idea in this paradigm was that AI revolves around the study of *cognition*. When we say that humans exhibit intelligence, we are not referring to their ability to recognize concepts, perceive objects, or execute complex motor skills, which they share with other animals. Rather, we mean they have the capacity to engage in multi-step reasoning, to understand the meaning of natural language, and to carry out problem solving in order to achieve novel goals. During AI's first 35 years, much of the discipline's research dealt with these issues, and the progress during that period arguably increased our understanding of the mind.

This belief is still active in some AI subfields, such as planning and constraint satisfaction, although each has developed its own specialized methods, but, unfortunately, other subareas have effectively abandoned their initial concern with cognition. For instance, machine learning focuses almost exclusively on classification and reactive control, whereas natural language processing has replaced its original emphasis on understanding with text classification and information retrieval. These shifts have produced short-term gains with many applications and clear performance improvements on their narrowly defined tasks. But I question whether advances on these fronts tell us much about the nature of cognition.

Another important assumption in early AI was that *knowledge* plays a central role in cognition, which in turn relies on the ability to represent and organize that knowledge. These claims depend on

the fundamental insight that computers are not simply number crunchers but rather general symbol manipulators. As Newell and Simon (1976) state clearly in their physical symbol system hypothesis, intelligent behavior requires the ability to interpret and manipulate symbolic list structures. The most impressive successes in AI's 50 year history, included the many examples of fielded expert systems, have relied on this capability.

Nevertheless, over the last ten years, many branches of AI have retreated from this position. The increased popularity of statistical and probabilistic methods has reduced the fragility of traditional symbolic schemes, but only at great losses in representational power. Some subfields, like machine learning and natural language processing, have almost entirely abandoned the goal of interpretable symbolic representations, caring only about performance, however achieved. This trend is very reminiscent of the behaviorist movement in psychology, which rejected the postulation of internal cognitive structures. Other subfields, like knowledge representation and constraint satisfaction, have retained a focus on symbols but limit the formalisms they consider for reasons of efficiency or analytical tractability. These developments constitute a major step back from the physical symbol system hypothesis, and they bode ill for our efforts to fathom the complex nature of intelligence.

Nowhere is this attitude more prevalent than in machine learning, a subfield in which I have been involved since its inception. Early work here dealt with the acquisition of symbolic cognitive structures, including logical concept definitions, recursive grammars, and symbolic heuristics for problem solving. There was a widespread assumption that the result of learning was declarative knowledge that had a clear interpretation and that would be used for reasoning, problem solving, or understanding. Machine learning initially aimed to support the acquisition of the full range of structures used in knowledge-based systems, as contrasted with the field of pattern recognition, which emphasized statistical methods for much more constrained tasks like classification.

In the late 1980s, a number of factors converged to change this situation, each a fine idea on its own but problematic when combined. One was the realization that machine learning should encompass all computational methods for improving performance from experience. This opened the door to ideas from pattern recognition like Bayesian classifiers, nearest neighbor methods, and neural networks. Another was the call to evaluate learning systems in terms of clear performance metrics like classification accuracy and problem-solving efficiency. The advent of the UCI repository of data sets made this increasingly easy for supervised classification learning, typically encoded as attribute-value pairs, which were well suited for variants on statistical pattern recognition.

In parallel, the early applications of machine learning technology took a similar path, focusing on supervised learning with attribute-value representations (Langley & Simon, 1995). The arrival of the data-mining movement in the mid-1990s demonstrated that many commercial problems fit well into this limited framework, and the subsequent rise of the World Wide Web encouraged rapid growth of similar work on learning from text. Both of these movements have been concerned primarily with improving predictive accuracy rather than with acquiring cognitive structures that support intelligent behavior, which was the original motivation for launching machine learning as a subfield of artificial intelligence.

Another central assumption of the cognitive systems paradigm was that intelligence involves *heuristic search* (Newell & Simon, 1976). Although not the only field to adopt the search metaphor,

AI was distinctive in its use of heuristics that, although not guaranteed to produce results, often make problems tractable which cannot be solved otherwise. On this dimension, AI differed from fields like operations research, which limited its attention to tasks for which one could find optimal solutions efficiently. Instead, many AI researchers had the audacity to tackle more difficult problems to which such techniques did not apply. Their approach involved developing search methods that relied on heuristic methods to guide search down promising avenues and that *satisficed* rather than found optimal solutions.

Unfortunately, the past decade has seen many AI researchers turn away from this practical attitude and adopt other fields' obsession with formal guarantees. For example, much recent work in knowledge representation has focused on constrained formalisms that promise efficient reasoning, even though this restricts the reasoning tasks they can handle. Research on reinforcement learning often limits itself to methods that provably converge to an optimal control policy, even if the time required for convergence makes them completely impractical. Also, the popularity of statistical approaches has resulted largely from the belief, often mistaken, that techniques with mathematical formulations provide guarantees about their behavior. One should certainly use nonheuristic methods when they apply to a problem, but it is another matter entirely to work only on tasks that such methods can handle. The original vision of AI was to address the same broad class of tasks as humans, but many now hope to redefine the field as something far more narrow.

This point relates to another assumption prevalent in early AI research – that the design and construction of intelligent systems has much to learn from the study of *human* cognition. Many central ideas in knowledge representation, planning, natural language, and learning (including the importance of heuristic search) were originally motivated by insights from cognitive psychology and linguistics, and many early, influential AI systems doubled as computational models of human behavior. The field also looked to human activities for likely problems that would challenge existing capabilities. Research on expert medical diagnosis, intelligent tutoring systems, artistic composition, and scientific discovery were all motivated by a desire to support activities considered difficult for humans.

Even in the first days of AI, few researchers attempted to model the details of human behavior, but they exhibited a genuine interest in psychology and the ideas it offered. But as time passed, fewer and fewer adopted this perspective, preferring instead to draw their inspirations and concerns from more formal fields. Still worse, fewer chose to work on challenging intellectual tasks that humans can handle only with considerable effort or advanced training. Attention moved instead to problems on which computers can excel using simple techniques combined with rapid computing and large memories, like data mining and information retrieval. There is no question that these efforts have had practical benefits, but they make no contact with psychology and they reveal little about the nature of intlligence in humans or machines.

Despite these changes, I believe the assumptions and methods of the cognitive systems paradigm remain as valid now as they were in the first days of AI. They hold our best hope for achieving the original goals of our field, they have been abandoned by the mainstream for insufficient reasons, and they deserve substantially more attention than they have received in recent years. If so, then

we should ask how we can resurrect interest in this approach to understanding intelligence and encourage its wider adoption within the AI community.

One important avenue concerns education. Most AI courses ignore the cognitive systems perspective, and few graduate students read papers that are not available on the Web, which means they are often unfamiliar with the older literature. Instead, we must provide a broad education in AI that cuts across different topics to cover all the field's branches and their role in intelligent systems. The curriculum should incorporate ideas from cognitive psychology, linguistics, and logic, which are far more important to the AI agenda than ones from mainstream computer science. One example comes from a course on reasoning and learning in cognitive systems (http://cll.stanford.edu/reason-learn05/), which I have offered (despite some internal opposition) through the Computer Science Department at Stanford University, but we need many more.

We should also encourage more research within the cognitive systems tradition. Funding agencies can have a substantial effect here, and the past few years have seen encouraging developments on this front, as DARPA IPTO has supported a number of large-scale programs with a cognitive systems emphasis (Brachman & Lemnios, 2002). I hope these projects will help excite both junior and senior researchers about the original vision of artificial intelligence. Of course, we also need venues to publish the results of such research, and there have been positive changes here as well. The annual AAAI conference now has a distinct track for integrated systems, and a new meeting on AI for interactive entertainment has a similar emphasis. We need more alternatives along these lines, but they are moves in the right direction.

We would also benefit from more audacious and visionary goals to spur the field toward greater efforts on cognitive systems. The General Game Playing competition (http://games.stanford.edu) is one promising development designed to foster research on general intelligent systems, and DARPA's plans for a 'cognitive decathlon', which would test abilities on a diverse set of cognitive tasks, is another good sign. But we also need programs that aim to demonstrate flexible, human-level behavior in everyday domains like in-city driving, which are constrained enough to be tractable yet rich enough to support the entire range of capabilities that we label as intelligent in people. The Turing test has many drawbacks but the right spirit; we need more efforts toward integrated systems that support the same breadth and flexibility as humans exhibit.

In summary, the original vision of AI was to understand the principles that support cognitive processing and to use them to construct computational systems with the same breadth of abilities as humans. As pursued within the cognitive systems paradigm, the field studied the content and representation of symbolic knowledge, the acquisition of such knowledge through learning, and the role of heuristic search in multi-step reasoning and problem solving. Human behavior provided a source of ideas for AI programs and many such systems served as models of this behavior. These ideas have lost none of their power or potential, and our field stands to benefit from their readoption by researchers and educators. Without them, AI seems likely to become a set of narrow, specialized subfields that have little to tell us about intelligence. Instead, we should use the assumptions of the cognitive systems approach as heuristics to direct our search toward true theories of the mind. This seems the only intelligent path.

References

Brachman, R., & Lemnios, Z. (2002). DARPA's new cognitive systems vision. *Computing Research News*, *14*, 1.

Langley, P., & Simon, H. A. (1995). Applications of machine learning and rule induction. *Communications of the ACM*, *38*, November, 55–64.

Newell, A., & Simon, H. A. (1976). Computer science as empirical enquiry: Symbols and search. *Communications of the ACM*, *19*, 113–126.