Artificial Intelligence and Cognitive Systems

PAT LANGLEY
Computing Science and Engineering
Arizona State University, Tempe, AZ 85287 USA

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 as though it were one of the core questions of science, on an equal footing with questions about 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, since its founding at the 1956 Dartmouth meeting, seemed impressive enough to promise rapid strides toward a broad 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 of them sociological, they have fallen into disfavor among many AI researchers. After this, I consider whether they have a role to play in the future 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*, a term championed by Brachman and Lemnios (2002).

High-Level Cognition

One key idea in this paradigm was that AI revolves around the study of *high-level 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 like dogs and cats. Rather, we mean that they have the capacity to engage in multi-step reasoning, to understand the meaning of natural language, to design innovative artifacts, to generate novel plans that achieve goals, and even to reason about their own reasoning. 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 idea is still active in some AI subfields, such as planning and automated reasoning, although each has developed its own specialized methods, but, unfortunately, other subareas have effectively abandoned their initial concern with high-level cognition. For instance, machine learning, despite its early interest in complex tasks, now 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 intelligence. A few researchers who take the cognitive systems perspective continue to address high-level behavior (e.g., Friedman, Forbus, & Sherin, 2011; Scally, Cassimatis, & Uchida, 2011), but we need far more work in this important area.

Structured Knowledge

Another important assumption in early AI was that *structured 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 – arguably the foundation of the 1956 AI revolution – that computers are not simply numeric calculators but rather general symbol manipulators. As Newell and Simon (1976) state clearly in their physical symbol system hypothesis, intelligent behavior appears to require the ability to interpret and manipulate symbolic structures. The most impressive successes in AI's 55 year history, including the many examples of fielded expert systems, have relied on this fundamental capability.

Nevertheless, over the last 20 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 techniques, but only at great losses in representational power. Some subfields have almost entirely abandoned the use of interpretable symbolic representations, caring only about performance, however achieved. This trend is 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. Some scientists continue to assume less constrained formalisms (e.g., Fahlman, 2011; Schubert, Gordon, Stratos, & Rubinoff, 2011), but we need more than a small cadre working in this arena. Such developments constitute a step backward from the physical symbol system hypothesis, and they distract from efforts to fathom the complex nature of intelligence.

Nowhere is this attitude more prevalent than in machine learning. Early work here dealt with the acquisition of symbolic cognitive structures, and there was a widespread assumption that mechanisms should produce easily interpreted declarative knowledge for use in reasoning, problem solving, or understanding. Machine learning initially aimed to support acquisition of the full range of structures used in knowledge-based systems, as contrasted with pattern recognition, which emphasized more constrained tasks like classification or categorization. As I have described elsewhere (Langley, 2011), in the late 1980s a number of factors converged to change this situation, including the influx of pattern-recognition techniques, the call for evaluation using metrics like classification accuracy, and the UCI repository's emphasis on attribute-value notations, which was well suited to statistical approaches. Early applications of machine learning (Langley & Simon, 1995) also focused on supervised learning with attribute-value notations, and the arrival of the data-mining movement and the World Wide Web in the mid-1990s demonstrated that many commercial problems fit this limited framework. Both emphasize induction of statistical predictors from large data sets, forgetting the original charter of machine learning was to acquire structured knowledge from limited experience.

System-Level Research

A third theme that characterized much early AI work was an emphasis on *system-level* accounts of intelligence. Because researchers envisioned comprehensive theories of the mind, they naturally recognized the need for their programs to comprise a number of interacting components. Many AI systems were given distinctive names that served as shorthand for a constellation of mutually supportive mechanisms. A related trend was the development of high-level programming languages, such as Prolog (Clocksin & Mellish, 1981), each with a distinctive syntax that reflected their theoretical assumptions about intelligence. These two ideas merged in Newell's (1990) notion of a *cognitive architecture*, which provided an infrastructure for building intelligent agents.

Despite these promising beginnings, by the 1990s many researchers had come to focus their energies on component algorithms rather than integrated systems. This resulted partly from AI finding its primary home in computer science departments, which gave higher status to the study of algorithms. Another influence was the emphasis on conference publications, which provided sufficient space to describe algorithms but not enough for system-level accounts. A third factor was the relative ease of evaluating algorithms, both formally and experimentally, which made it easier to produce and publish papers on such topics. Finally, university professors found it far simpler to teach AI as a set of unrelated algorithms than to present coherent frameworks for intelligent systems. The results have been a greatly decreased interest in system-level accounts and the fragmentation of AI into a set of disconnected subfields. Research on cognitive architectures (Langley, Laird, & Rogers, 2009) provides some important counterexamples to this trend, but it is in the minority and intelligent systems deserve far more attention.

Heuristics and Satisficing

Another central assumption of initial AI research was that intelligence involves *heuristic search* (Newell & Simon, 1976). Although not the only field to adopt the search metaphor, it was distinctive in its use of heuristics that, although not guaranteed to produce results, often made problems tractable which could not be solved otherwise. On this dimension, AI differed from other fields, such as operations research, that limited their 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 heuristics to guide search down promising avenues and that *satisficed* (Simon, 1956) by finding acceptable rather than 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 address. 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. We should certainly use nonheuristic methods when they apply to a problem, but this does not mean we should only study tasks that such techniques can handle. Of course, some work on heuristic approaches continues (e.g., Bridewell & Langley, 2011; MacLellan, 2011), but it is rare and often held in low regard by acolytes of the AI mainstream. The original charter 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.

Links to Human Cognition

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 (e.g., Newell, Shaw, & Simon, 1959). The field also looked to human activities for likely problems that would challenge existing capabilities. Research on expert medical diagnosis (Shortliffe & Buchanan, 1975), intelligent tutoring systems (Sleeman & Brown, 1982), artistic composition (Cohen, 1975), and scientific discovery (Langley, 1981) were all motivated by a desire to support activities considered difficult for humans. In fact, the title of this publication reflects the early association between the two disciplines.

Even in the first days of AI, few researchers attempted to model the details of human behavior, but many exhibited a genuine interest in psychology and in 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.[1] 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 intelligence in humans or machines. Again, some researchers continue to make draw upon results about human cognition, but such efforts are few and far between.

The Future of Cognitive Systems

Despite these changes, I believe the assumptions and methods of the cognitive systems paradigm remain as valid now as they were over 50 years ago, 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 research 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 artificial intelligence and cognitive systems (http://circas.asu.edu/aicogsys/) that I have offered at Arizona State University, but we need many more.

1. Even for challenging problems like playing chess that require heuristic search, the vast majority of work has come to rely heavily on fast CPUs and large storage.

We should also encourage more research within the cognitive systems tradition. Funding agencies can have a major effect here, and the past decade has seen encouraging developments on this front. During this period, DARPA in the USA supported a number of large-scale programs with a cognitive systems emphasis (Brachman & Lemnios, 2002), and the US Office of Naval Research has long shown a commitment to the paradigm. The European Union has also funded substantial projects (e.g., Christiansen, Sloman, Krujiff, & Wyatt, 2009) in the area.[2] Continued government support of cognitive systems research will aid progress, but we need committed people to join funding agencies as program officers to ensure that this occurs.

The field would also benefit from more audacious and visionary goals to spur the field toward greater efforts on cognitive systems. For instance, the General Game Playing competition (http://games.stanford.edu) has fostered research on general intelligent systems, and proposals for a 'cognitive decathlon' that would measure abilities on a set of well-defined cognitive tests is another good sign. But we also need demonstrations of flexible, high-level cognition in less constrained settings that require the combination of inference, problem solving, and language into more complete intelligent systems. The Turing test has many drawbacks but the right spirit, and we need more efforts toward integrated systems that support the same breadth and flexibility as humans. Challenging tasks 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 research on cognitive systems. From 2006 to 2011, the annual AAAI conference included a special track on 'integrated intelligence' that encouraged submissions on system-level results. The recent AAAI Fall Symposia on Advances in Cognitive Systems (http://www.cogsys.org/acs/2011/) attracted over 75 participants, and its organizers plan to launch a regular conference during 2012, along with an associated electronic journal. We need more alternatives along these lines to help counter the mainstream bias in favor of papers that report on narrow tasks, standalone algorithms, and incremental performance improvements. Broader criteria for scientific progress are necessary to advance the field, making room for papers that analyze challenging problems, demonstrate new functionalities, and replicate capabilities that are distinctively human.

In summary, the original vision of AI was to understand the principles that support high-level 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. Much of this research focused on integrated systems rather than component algorithms, and cognitive psychology provided a source of ideas for these programs, many of which served as models of human 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.

2. Not all work funded under these programs, in either the US or Europe, has focused on cognitive systems as we have defined them, but even researchers who hold views antithetical to those reviewed here are sometimes attracted to the movement's higher-level theme.

References

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

Bridewell, W., & Langley, P. (2011). A computational account of everyday abductive inference. *Proceedings of the Thirty-Third Annual Meeting of the Cognitive Science Society*. Boston.

Christiansen, H. I., Sloman, A., Krujiff, G.-J., & Wyatt, J. (2009). *Cognitive systems*. Berlin: Springer-Verlag.

Clocksin, W. F., & Mellish, C. S. (1981). *Programming in Prolog*. Berlin: Springer-Verlag.

Cohen, H. (1975). Getting a clear picture. *Bulletin of the American Society for Information Science*, December, 10–12.

Fahlman, S. E. (2011). Using Scone's multiple-context mechanism to emulate human-like reasoning. *Advances in Cognitive Systems: Papers from the 2011 AAAI Fall Symposium* (pp. 98–105). Arlington, VA: AAAI Press.

Friedman, S., Forbus, K. D., & Sherin, B. (2011). Constructing and revising commonsense science explanations: A metareasoning approach. *Advances in Cognitive Systems: Papers from the 2011 AAAI Fall Symposium* (pp. 121–128). Arlington, VA: AAAI Press.

Langley, P. (1981). Data-driven discovery of physical laws. *Cognitive Science*, *5*, 31–54.

Langley, P. (2011). The changing science of machine learning. *Machine Learning*, *82*, 275–279.

Langley, P., Laird, J. E., & Rogers, S. (2009). Cognitive architectures: Research issues and challenges. *Cognitive Systems Research*, *10*, 141–160.

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

MacLellan, C. J. (2011). An elaboration account of insight. *Advances in Cognitive Systems: Papers from the 2011 AAAI Fall Symposium* (pp. 194–201). Arlington, VA: AAAI Press.

Newell, A., Shaw, J. C., & Simon, H. A. (1959). Report on a general problem-solving program. *Proceedings of the International Conference on Information Processing* (pp. 256-264). Paris: UNESCO House.

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

Scally, J. R., Cassimatis, N. L., & Uchida, H. (2011). Worlds as a unifying element of knowledge representation. *Advances in Cognitive Systems: Papers from the 2011 AAAI Fall Symposium* (pp. 280-287). Arlington, VA: AAAI Press.

Schubert, L. K., Gordon, J., Stratos, K., Rubinoff, A. (2011). Towards adequate knowledge and natural inference. *Advances in Cognitive Systems: Papers from the 2011 AAAI Fall Symposium* (pp. 288-296). Arlington, VA: AAAI Press.

Shortliffe, E. H., & Buchanan, B. G. (1975). A model of inexact reasoning in medicine. *Mathematical Biosciences*, *23*, 351–379.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*, 129–138.

Sleeman, D., & Brown, J. S. (Eds.) (1982). *Intelligent tutoring systems*. New York: Academic Press.