Artificial Intelligence

J Feldman

This article is reproduced from 'Artificial Intelligence in Cognitive Science' in the *International Encyclopedia of the Social & Behavioral Sciences*, Vol 2, pp. 792–796, © 2001, Elsevier Science Ltd.

Overview

Artificial intelligence (AI) and cognitive science are two distinct disciplines, with overlapping methodologies but with rather different goals. AI is a branch of computer science and is concerned with construction and deployment of intelligent agents as computer programs, and also with understanding the behavior of these artifacts. The core scientific goal of AI is to understand the basic principles of intelligent behavior that apply equally to animal and artificial systems. Almost all of the work is mathematical or computational in character and much of the literature is technique oriented.

Cognitive science is an explicitly interdisciplinary field that has participation not only from AI, but also from linguistics, philosophy, psychology, and subfields of other social and biological sciences. The unifying goal of cognitive science is to understand and model human intelligence, using the full range of findings and methodologies of the complementary disciplines. As one would expect, a wide range of techniques from the mathematical, behavioral, social, and biological sciences are employed. Cognitive science, in contrast with AI, is defined more by phenomena than by methodology. There are research groups that are active in both AI and cognitive science, but they tend to produce different types of reports for journals and conferences in the two areas.

Shared Origins in the Postwar Cognitive Revolution

Both AI and cognitive science evolved after 1950, and in their early development were more tightly integrated than at present. For much of the first half of the twentieth century, the Anglo-American study of cognition was dominated by the behaviorist paradigm, which rejected any investigation of internal mechanisms of mind. The emergence of both AI and cognitive science was part of a general postwar movement beyond behaviorist theories, which also included new approaches in linguistics and the social sciences. The idea of computational models of mind was a central theme of what is sometimes called the 'postwar cognitive revolution.' One of the leading early AI groups, under the leadership of Allen Newell and Herbert Simon at Carnegie Mellon, was explicitly concerned with cognitive modeling using the symbolic processes of AI. The current version of this continuing effort is the symbolic cognitive architecture known as Soar (the term derived originally from the concept 'state, operator, and result,' but those currently involved in Soar development prefer not to represent the term as an acronym). Another traditional symbolic approach to modeling intelligence is known as ACT (an acronym for 'adaptive control of thought'). However, there is very little work in contemporary AI that is explicitly focused on modeling human behavior as opposed to intelligent systems in general. There is some continuing work on human and machine game playing, but it is not integrated into the fields of AI and cognitive science.

The central idea of AI is computational modeling of intelligent behavior – its main contribution to cognitive science. The basic notion of a computational model is now commonplace in all scientific fields and many other aspects of contemporary life. One builds a detailed software model of some phenomenon and studies the behavior of the model, hoping to gain understanding of the original system. Much of the work in AI has the engineering goal of producing practical systems, and there is no sharp boundary between AI and other applied fields of computer science and engineering. AI techniques are now commonplace in the full range of business, scientific, and public applications. While all fields use computational models, researchers, in computer science in general and AI in particular, invent and study computational techniques for constructing models, presenting the results of simulations, and understanding the limitations of the simulation. AI has traditionally studied the modeling of the most complex phenomena – those relating to intelligence. Because of the technical challenges arising in the construction of these complex simulations, many innovations in computing have arisen in AI and then have been more widely applied.

Domain-Focused Research That Cuts Across AI and Cognitive Science

The relationship between AI and cognitive science is further complicated by the fact that there are currently several distinct research fields that cut across both disciplines, but have separate journals, meetings, etc. The most prominent of these research areas are speech, language, vision, and neural networks. Each of these fields has thousands of practitioners, many of whom are interested in AI, cognitive science, or both. Appropriately, each of these areas is represented by a vast body of literature. As AI and cognitive science have grown, specialized areas such as language and vision modeling have become largely independent, but they do continue to share the development of underlying methodologies. The main areas that have remained as core AI include knowledge representation and reasoning, planning, and problem solving. The study of learning has evolved somewhat differently (see later). There are some common scientific paradigms that cut across all of these fields, and these are discussed in the following sections as they relate to the social and behavioral sciences. The common thread linking AI to cognitive science is reliance on computational models of different kinds.

Role of Formal Logic in AI and Cognitive Science

To a great extent, the early development of AI was based on symbolic, as opposed to numerical, modeling. This led to the introduction of some novel representations such as those of Soar, but the main effect of this was to align AI with formal logic for much of its early history. In fact, much of the driving force for the creation of the field called cognitive science came from people who saw mathematical logic as its unifying theme. This remains a fruitful approach to AI and constitutes one major area of overlap with cognitive science. Mathematical logic is elegant and well developed and can be shown to be, in some sense, general enough to represent anything that can be described formally. There remains a significant community of linguists, philosophers, and computer scientists for whom logic is the only scientific way to study intelligence.

However, the twentieth century was not kind to categorical, deterministic theories in any field, and cognitive science was no exception. A central issue in the study of cognition has been how to describe the meaning of words and concepts. In formal logic, a concept is defined by a set of necessary and sufficient conditions. For example, a bachelor might be defined as a male who never married. The limitations of classical, all-or-none, categories were already recognized by Wittgenstein, who famously showed that concepts such as 'game' could not be characterized by necessary and sufficient conditions, and were better described as family resemblances. Even the definition of 'bachelor' becomes graded when we consider cohabitation, or how old a male needs to be before being considered a bachelor.

Starting in the 1960s, a wide range of cognitive science studies by Rosch and others showed the depth and complexity of human conceptual systems and their relation to language. This helped give rise to the subfield called cognitive linguistics, which overlaps with cognitive science, but has its own paradigms, journals and conferences. The graded and relational nature of human categories undermined the attempt to create a unified science of mind based on formal logic. Another attack on the formalist program arose from the growing understanding of the neural basis of intelligence. A crucial insight was that basic human concepts are grounded in direct experience and that more abstract concepts are mapped metaphorically to more embodied ones. This undercuts the formalist program, and also leads to a separation of AI, which studies intelligence in the abstract, from cognitive science, which is explicitly concerned with human minds. While much excellent work continues to be done using logic, it is now generally recognized that there are a wide range of phenomena that are better handled by biologically based models and/or some form of numerical, often probabilistic, modeling. For a variety of reasons, the movement to quantitative numerical models followed somewhat different paths in AI and cognitive science, but there are some recent signs of reconvergence.

Learning and the Connectionist Approach to Cognitive Science

As it happens, while the formalists were trying to establish a cognitive science based on formal logic, an antithetical neural network movement was also developing, and this approach has become a major force in cognitive science. The two contrasting approaches to cognitive modeling, neural modeling and logic, were mirrored in the two different methods by which early computer scientists sought to achieve AI. From the time of the first electronic computers around 1950, people dreamed of making them 'intelligent' by two quite distinct routes. The first, 'conventional' AI, is to build standard computer programs as models of intelligence. This remains the dominant paradigm in AI and has had considerable success. The other approach is to try to build hardware that is as brainlike as possible and have it learn the required behavior. The history of this 'neural modeling' approach has been well described. After some promising early mathematical results on learning in simple networks, the neural learning approach to modeling intelligence fared much less well for three decades and had little scientific or applied success. Around 1980, a variety of ideas from biology, physics, psychology, and computer science and engineering coalesced to yield a 'new connectionist' approach to modeling intelligence that has become a core field of cognitive science, and also the basis for a wide range of practical applications. Among the key advances was a mathematical technique (back-propagation) that extended the early results on learning to a much richer set of network structures.

Connectionist computational models are almost always computer programs, but programs of a different kind than those used in, for example, word processing or symbolic AI. Connectionist models are specified as a network of simple computing units, which are abstract models of neurons. Typically, a model unit calculates the weighted sum of its inputs from upstream units and sends to its downstream neighbors an output signal that is a nonlinear function of its inputs. Learning in such systems is modeled by experience-based changes in the weights of the connections between units. The basic connectionist style of modeling is now being used in three quite different ways – in neurobiology, in applications, and in cognitive science. Neurobiologists who study networks of neurons employ a wide range of computational models, from very detailed descriptions of the internal chemistry of the neuron to the abstract units just described. The use of connectionist neural models in practical applications is part of the reconvergence with AI and is discussed in the final section of this article.

In cognitive science, connectionist techniques have been used for modeling all aspects of language, perception, motor control, memory, and reasoning. This universal coverage represents a potential breakthrough; previously, the computational models of, for example, early vision and problem solving used entirely different mathematical and computational techniques. Since the brain is known to use the same neural computation throughout, it is not surprising that neurally inspired models can be applied to all behavior. Unfortunately, the existing models are neither broad nor deep enough to ensure that the current set of mechanisms will suffice to bridge the gap between structure and behavior, but the work remains productive.

Connectionist models in cognitive science fall into two general categories, often called structured and layered networks (also called parallel distributed processor, or PDP, networks). Most modelers are primarily interested in learning, which is modeled as experience-driven change in connection weights. There is a great deal of research studying different models of learning with and without supervision, different rules for changing weights, etc. Because of the focus on what the network can learn, any prewired structure will weaken the results of the experiment. The standard approach is to use networks with unidirectional connections arranged in completely connected layers, sometimes with a very restricted additional set of feedback links. This kind of network contains a minimum of presupposed structure and is also amenable to efficient learning techniques, such as the aforementioned back-propagation method. Most researchers using totally connected layered models do not believe that the brain shares this architecture, but there is an ongoing controversy about the implications of PDP learning models for theories of mind (see the section titled 'Nature and nurture: rules versus connections').

Structured connectionist models are usually less focused on learning than on the representation and processing of information. Essentially all the modeling done by neurobiologists involves specific architectures, which are known from experiment. For structured connectionist models of cognitive phenomena, the underlying brain architecture is rarely known in detail and sometimes not at all at the level of neurons and connections. The methodology employed is to experiment with computational models of the behavior under study that are consistent with the known biological and psychological data and are also plausible in the resources (neurons, computing time, etc.) required. This methodology is very similar to what are called 'spreading activation' models, widely used in psycholinguistics. Some studies combine structured and layered networks, or investigate learning in networks with an initial structure that is tuned to the problem area or the known neural architecture.

Nature and Nurture: Rules versus Connections

Perhaps the most visible contribution to date of connectionist computational models in cognitive science has been to provide a new jousting ground for contesting some age-old issues on the nature of intelligence. Much of the debate has been published in *Science* magazine, which suggests that it is considered to be of major importance by the US scientific establishment. The nature versus nurture question concerns how much of some trait, usually intelligence, can be accounted for by genetic factors, and how much depends on postnatal environment and training. Some PDP connectionists have taken very strong positions, suggesting that learning can account for everything interesting. In the particular case of grammar, an important group of linguists and other cognitive scientists take an equally extreme nativist position, suggesting that humans only need to choose a few parameters to learn grammar. A related issue is whether human grammatical knowledge is represented as general rules or just appears as the rulelike consequences of PDP learning in the neural network of the brain. There is ample evidence against both extreme positions, but the debate continues to motivate a great deal of thought and experiment.

Current and Future Trends

Although the fundamental split between the AI focus on general methods and the cognitive science emphasis on human intelligence remains, there are a growing number of areas of overlapping interest. As already discussed, quantitative neural models are playing a major role in cognitive science. It turns out that the mathematical and computational ideas underlying learning in neural networks have found application in a wide range of practical problems, from speech recognition to financial prediction. The basic idea is that, given current computing power, back-propagation and similar techniques allow large systems of nonlinear units to learn quite complex probabilistic relationships using labeled data. This general methodology overlaps not only with AI but also with mathematical statistics, and is part of a unifying area called computational learning theory. There is also a large community of scientists and engineers who identify themselves as working on neural networks and related statistical techniques for various scientific and applied tasks, along with conferences and journals to support this effort.

While probability was entering cognitive science from the bottom-up through neural models, AI experienced the introduction of probabilistic methods from general theoretical considerations, which only later led to practical application. As discussed earlier, the limitations of formal logic became well recognized in the 1960s. Over the subsequent decades, AI researchers, led by Judea Pearl, at the University of California (Los Angeles), developed methods for specifying and solving large systems of conditional probabilities. These belief networks are now widely used in applications ranging from medical diagnosis to business planning. A growing field that involves both symbolic and statistical techniques is 'data mining,' processing large historical databases to search for relationships of commercial or social importance. Recent efforts to learn or refine belief networks from labeled data are another area of convergence of AI and cognitive science in computational learning theory. Of course, the explosion of Internet activity is affecting AI along with the rest of the computing field. Two AI application areas that seem particularly important to cognitive science are intelligent Web agents and spoken-language interaction.

As the range of users and activities on the Internet continues to expand, there is increasing demand for systems that are both more powerful and easier to use. This is leading to increasing efforts on the human–computer interface, including the modeling of user plans and intentions – clearly overlapping with traditional concerns of cognitive science. One particularly active area is interaction with systems using ordinary language. Whereas machine recognition of individual words is relatively successful, dealing with the full richness of language is one of the most exciting challenges at the interface between AI and cognitive science, and a problem of great commercial and social importance.

Looking ahead, we can be confident that the increasing emphasis on intelligent systems will continue. From the scientific perspective, it is very likely that most of the interdisciplinary research in cognitive science will remain focused on specialized domains such as language, speech, and vision. General issues including representation, inference, and learning will continue to be of interest and will constitute the core of the direct interaction between AI and cognitive science. With the rapid advances in neurobiology, both fields will increasingly articulate with the life sciences, with great mutual benefits.

See also: Animal Intelligence: The Search for Animal Intelligence; Cognitive Neuroscience: An Overview; Cognitive Control and Development; Connectionist Models; Connectionist Models of Language Processing; Executive Function and Higher-Order Cognition: Computational Models; Hippocampus: Computational Models; Memory: Computational Models; Numerical Intelligence: Neural Substrates.

Further Reading

Elman J, Bates E, and Johnson M (1996) *Rethinking Innateness: A Connectionist Perspective on Development (Neural Network Modeling and Connectionism)*. Cambridge, MA: MIT Press.

Lakoff G (1987) *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind*. Chicago, IL: University of Chicago Press.

McClelland JL and Rumelhart DE (1986) *Parallel Distributed Processing*. Cambridge, MA: MIT Press.

Newell A (1990) *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.

Pearl J (1988) *Probabilistic Reasoning in Intelligent Systems*. San Mateo, CA: Morgan Kaufmann.

Posner MI (ed.) (1985) *Foundations of Cognitive Science*. Cambridge, MA: MIT Press.

Regier T (1996) *The Human Semantic Potential*. Cambridge, MA: MIT Press.

Russell SJ and Norvig P (2000) *Artificial Intelligence*. Upper Saddle River, NJ: Prentice-Hall.