Looking back, looking ahead: Symbolic versus connectionist AI

Ashok K. Goel

Georgia Institute of Technology, Atlanta, GA, USA

Correspondence
Ashok K. Goel, Georgia Institute of Technology Atlanta 30332, GA.
Email: ashok.goel@cc.gatech.edu

Abstract
The ongoing debate between symbolic and connectionist AI attends to some of the most fundamental issues in the field. In this column, I briefly review the evolution of the unfolding discussion. I also point out that there is a lot more to intelligence than the symbolic and connectionist views of AI.

Like much of the AI community, I have watched the ongoing discussion between symbolic AI and connectionist AI with fascination. While symbolic AI posits the use of knowledge in reasoning and learning as critical to producing intelligent behavior, connectionist AI postulates that learning of associations from data (with little or no prior knowledge) is crucial for understanding behavior. The recent debate between the two AI paradigms has been prompted by advances in connectionist AI since the turn of the century that have significant applications. The technological successes of connectionism in the presence of large-scale data have made it the dominant paradigm in AI. The conversation between the two schools has unfolded over the last decade through scholarly articles (for example, LeCun, Bengio, and Hinton, 2015), debates (AI Debate, 2017; AI Debate, 2019), and social media – with the last mentioned sometimes inviting sharp commentaries. My fascination with the debate has been because of the importance of the main issue to AI: the nature of intelligence itself. Yet, I have also found the debate to be a little frustrating. Here is why.

First, I find the phrases "symbolic AI" and "connectionist AI" misleading. The commitment of the symbolic school is to knowledge and its use in reasoning and learning (with only modest input data), not to symbols as such: symbols often have only stood for knowledge abstractions. Similarly, the allegiance of the connectionist camp really is to learning associations from data with little or no prior knowledge, typically requiring large-scale data. From the perspective of cognitive science, symbolic AI is well aligned with the rationalist school of mind with its empha-

sis on the use of knowledge acquired through biological evolution and cognitive development. Similarly, connectionist AI is related to the empiricist school of mind with its focus on the use of data acquired through sensory experiences in life (though connectionist AI is not limited to sensory data). Thus, the differences between the two schools are not only technological but also philosophical and the philosophical differences have a deep history. However, the terms symbolic AI and connectionist AI are so common that I will adhere to them here with the understanding that we are really talking about the rationalist and the empiricist schools.

Second, both camps tend to create and attack caricatures of the other. For example, symbolicists sometimes criticize specific connectionist architectures or algorithms such as the backpropagation algorithm in artificial neural networks. However, the connectionist's allegiance is to learning associations from data as the basis of intelligence, not to any particular algorithm or architecture. Similarly, the connectionists sometimes attack specific symbolic architectures or algorithms such as production systems on which the expert systems of the 1980s were based. But again the commitment of symbolic AI is to intelligence based on knowledge and inferencing, not to any specific representation or architecture. Thus, many of the critiques from both sides often are high on rhetoric but lacking in substance.

Third, the two sides often insist on interpreting the same thing very differently. For example, connectionist AI may produce an artificial neural network for an image recognition task and claim that the network does not contain

any prior knowledge or perform any reasoning over the knowledge. Symbolists may look at the same network and see prior knowledge in the form of network engineering (the design of the structure and the dynamics of the network), feature engineering (features of data that are input into the network), and sometimes even concept engineering (abstractions directly represented among the hidden units of the network). In fact, the successes of modern artificial neural networks such as convoluted neural networks arise from smart – indeed, beautiful – systematizing of network, feature, and concept engineering for achieving specific computational properties such as translational invariance (the ability to recognize an object in an image even if the object is shifted from one point to another). Sometimes the artificial neural network is explicitly complemented by symbolic machinery such as tree-based search in the famous AlphaGo program (Silver et al., 2016). Arguments about the exact role of this kind of knowledge, and whether the real source of power lies in the connectionist techniques or the symbolic structures, often lead to much heat but little insight.

Fourth, many of the arguments between symbolic AI and connectionist AI are repetitions from the 1980s. I was a PhD student at The Ohio State University when Rumelhart, McClelland and the PDP Group (1986) started publishing their three volume series on Parallel Distributed Processing (PDP). I recall the excitement in the AI research community about the potential for understanding and building intelligence in the empiricist school without requiring knowledge and inferencing. Then, as now, there was joy in large parts of the AI research community, and perhaps also a little surprise, that the connectionist techniques had been successful at an increasing number of tasks. Then, as now, we heard claims from connectionists that symbolic AI has failed, and that connectionist AI can do everything symbolic AI can do, or soon will, all without requiring knowledge and inferencing. Then, as now, we read about the skepticism of symbolists about some of the connectionist claims, and doubts that the connectionist models, even if successful at some narrow tasks, are actually intelligent in any deep sense.

Fifth, Marr's (1982) classical framework of three information-processing levels for understanding and designing intelligence provides one potential avenue for moving beyond the debate between symbolic and connectionist AI. The highest and most important level in Marr's framework describes how knowledge partitions a complex task into smaller, simpler subtasks so that the subtasks can be performed efficiently (Marr called this the "computational theory" of addressing the task); the middle level describes the representations and algorithms for addressing the various subtasks and thus accomplishing the complex task; and the third level at the bottom

pertains to the implementation of the algorithms in hardware or software. In 1988, my PhD advisor Balakrishnan Chandrasekaran, fellow graduate student Dean Allemang, and I wrote an article for the AI Magazine presenting an analysis of the symbolic vs connectionist AI debate in Marr's framework (Chandrasekaran, Goel, and Allemang, 1988). The article suggested that while symbolic AI and connectionist AI offer two different sets of abstractions and mechanisms for realizing a computational theory for addressing a complex task, the real and the hard action was at the level of building the computational theory itself. However, while most symbolists are willing to accept Marr's framework as a basis for understanding and designing intelligence, many connectionists are not

Sixth, the connectionist school posits that the implementational hardware at the bottom level in Marr's framework provides affordances and imposes constraints on the algorithms and representations in the middle, and that these bottom-up affordances and constraints can result in a very different set of algorithms for addressing the high-level task. This argument is quite valid. However, the connectionist school further postulates that it can exploit the affordances of parallel distributed hardware and large-scale data to build architectures and algorithms for accomplishing complex tasks without the benefits of symbolic representations such as reference, variable binding, type-token distinction, and modularity and compositionality, thereby negating the physical symbol system hypothesis (Newell and Simon, 1976). Although recently connectionist AI has started addressing problems beyond narrowly defined recognition and classification tasks, this mostly remains a promise: it remains to be seen if connectionist AI can accomplish complex tasks that require common-sense reasoning and causal reasoning, all without including knowledge and symbols.

Seventh, the similarity between the arguments in the early 2020s and the late 1980s goes even further than outlined above. Now, as then, there are calls for new techniques for explaining the processing in artificial neural networks because the processing often is opaque, for infusing knowledge into connectionist architectures to enable multi-step inferencing, for a third way of neuro-symbolic architectures that combine the advantages of the two paradigms. Now, as then, most of these calls are coming from symbolists, perhaps because they help make their point about the need for knowledge and inferencing. Recent theories in cognitive science that propose dual processes for producing human behavior – sometimes called System 1 and System 2 (Kahneman, 2011; Stanovich and West, 2000) – provide a theoretical framework for reconciling aspects of symbolic AI and connectionist AI. According to the dual-process theories of mind, System 1 is associative, tacit, imagistic, personalized, and fast, while System

2 is analytical, explicit, verbal, generalized, and slow. However, it is important to note that the mapping between symbolic AI and connectionist AI on one hand and System 1 and System 2 in human cognition on the other is not a direct one-to-one mapping. While System 1 likely contains abstractions and algorithms of both symbolic and connectionist AI varieties, the abstractions and algorithms of System 2 likely are mostly symbolic (though of course it too is implemented on neural networks in the human brain).

Eighth, it is important and useful to remember that there is a lot more to the nature of intelligence than the debate between symbolic AI and connectionist AI. Over the last 30 years, cognitive science has expanded its view of mind to include embodied cognition, situated cognition, distributed cognition, and social and cultural cognition, all of whom place significant parts of mind outside an individual human's head. However, the same kind of expansion of scope has not yet occurred in AI. Recently I have become enamored of theories of socially situated cognition according to which human learning is fundamentally a social process. We learn by observing other humans, by emulating and imitating our parents, teachers, models, and mentors. We have developed social structures, such as families, schools, temples, and laboratories, for teaching and learning through instruction, demonstration, exploration, and collaboration. Yet, neither symbolic nor connectionist AI have much to say about socially situated intelligence.

This provides an opportunity for both symbolic and connectionist AI, and especially, I think, for symbolic AI: develop new computational theories of socially situated intelligence (as well as embodied intelligence, physically situated intelligence, distributed intelligence, and social and cultural intelligence) that place significant parts of a machine's "mind" outside its "head." As just one example, there is much to be done in the space of designing intelligent agents that can learn from and teach other intelligent agents including humans, that can use interactions with humans to develop a mutual theory of mind, and that can foster better human-human communication and collaboration.

REFERENCES

AI Debate: Y. Bengio vs. G. Marcus. 2019. MILA AI Institute. https://www.youtube.com/watch?v=EeqwFjqFvJA (last retrieved on April 17, 2021).

AI Debate: Y. LeCun vs. G. Marcus. Does AI Need More Innate Machinery? 2017. NYU Center for Mind, Brain and Consciousness. https://www.youtube.com/watch?v=aCCotxqxFsk (last retrieved on April 17, 2021).

Chandrasekaran, Balakrishnan, Ashok Goel, and Dean Allemang. 1988. "Connectionism and information processing abstractions: The message still counts more than the medium." *AI Magazine* 9(4): 24–34.

Kahneman, Daniel. 2011. *Thinking Fast and Slow*. New York: MacMillan.

LeCun Yann, Yoshua Bengio, and Geoffrey Hinton. 2015. "Deep learning." *Nature* 521: 436–44.

Marr, David. 1982. *Vision: A Computational Approach*. San Francisco: Freeman & Co.

Newell, Allen, and Herbert A. Simon. 1976. "Computer science as empirical inquiry: Symbols and search." *Communications of the ACM* 19(3): 113–26.

Rumelhart, David E., James L. McClelland, and the PDP Group. 1986 *Parallel and Distributed Processing: Explorations in the Microstructure of Cognition. Volumes I and II*. Cambridge, MA: MIT Press.

Silver, David, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, et al. 2016. "Mastering the game of Go with deep neural networks and tree search." *Nature* 529: 484–89.

Stanovich, Keith E., and Richard F. West. 2000. "Individual differences in reasoning: Implications for the rationality debate." *Behavioral and Brain Sciences* 23: 645–726.

AUTHOR BIOGRAPHY

Ashok Goel is a Professor in the School of Interactive Computing at Georgia Institute of Technology and the Chief Scientist with Georgia Tech's Center for 21st Century Universities. He is a Fellow of AAAI, a recipient of AAAI's Outstanding AI Educator Award, and an Editor Emeritus of AI Magazine. He is the Executive Director of the new NSF National AI Research Institute on Adult Learning and Online Education.