

AI's Half-Century¹

Margaret A. Boden

■ The first 50 years of AI are reviewed, and current controversies outlined. Scientific disputes include disagreements over the best research methodology, including classical AI, connectionism, hybrid systems, and situated and evolutionary robotics. Philosophical disputes concern (for instance) whether computation is necessary and sufficient for mentality, whether representations are essential for intelligence, whether consciousness can be explained objectively, and whether the Cartesian presuppositions of (most) AI should be replaced by a neo-Heideggerian approach. With respect to final verdicts, both juries (scientific and philosophical) are still out. But AI has aided theoretical psychology and revived the philosophy of mind.

Astrologers would have a hard time with AI, for it's difficult to say just when this particular baby was born. As good a date as any, however, is 1943—almost exactly half a century ago.

In that year, Warren McCulloch (a psychiatrist, cybernetician, philosopher, and poet) and Walter Pitts (a research student in mathematics) published a seminal paper combining early twentieth-century ideas on computation, logic, and the nervous system. The result was a heady brew, which explicitly promised to revolutionize psychology and philosophy—and which, in the event, revolutionized technology too.

How AI Began

McCulloch and Pitts' paper ("A Logical Calculus of the Ideas Immanent in Nervous Activity") concentrated on how propositions expressible in logic could be computed by simple neural nets. Those nets consisted of cells passing inhibitory and excitatory messages between them and acting as what computer scientists (soon afterwards) called "and-

gates," "or-gates," and the like. The logical design of the first digital computer, in the hands of John Von Neumann, was strongly influenced by this work of 1943.

Von Neumann himself rejected logic as the key to human thinking. Something much more like thermodynamics, he speculated, is needed. And in 1947—only just under half a century ago—Pitts and McCulloch published "How We Know Universals: The Perception of Auditory and Visual Forms." This described the brain as a parallel-processing, self-equilibrating system, changing according to statistical equations like those used in physics.

Their first paper made many intellectual waves—which are still spreading, 50 years later. They had claimed that the truth or falsity of any (computable) proposition could, in principle, be computed by a simple type of neural net. The future of psychology, they said, consisted of the design of various sorts of neural networks (logical circuits). This novel methodology, and the nascent technology associated with it, promised to show just how mind is grounded in mechanism.

By the 1950s, computer-based work on the simulation of mental processes had already started. Much of this was "logical" in nature and developed into what's known as classical, or symbolic, AI. But some was what is nowadays called *connectionist*, studying networks of simple computational units, communicating by excitatory and inhibitory links. Connectionism went into relative decline during the late 1960s. In the late 1980s, however, it blossomed—hitting the news-stands with rash promises of "brainlike" computers just around the corner. But both these forms of AI share the same historical roots.

So much for pedigree. But does a mere half-century of work count as a pedigree? Might it rather be a mere blip, an unfortunate academic mutation with no real intellectual fitness?

And does pedigree confer respectability? If respectability implies noncontroversiality, then AI—of whatever type—is not yet respectable. On the contrary, it is highly disputatious, both scientifically and philosophically.

Current Scientific Disputes within AI

AI scientists themselves favor differing research methodologies. To some extent, this depends on what problems they're interested in. *This* methodology may be appropriate for one type of problem, *that* for another. But most AI researchers prefer to work within a particular computational "paradigm."

Classical AI is still a leading contender. It's widely used in commercial AI. Applications of classical AI include natural language interfaces to programs of many kinds as well as expert systems used by various industrial, financial, medical, military, and other public institutions. It's widely used for research purposes, too, in the study of problem solving, planning, learning, natural language, analogy, the perception and performance of music, and creativity in art and science. Some people are using it to explore the psychology of motivation and emotion too.

The advantages of classical AI include its abilities to represent hierarchical structure, to define "strong" (exceptionless) problem constraints, and to provide models whose functioning is relatively easy to understand.

Another commonly used approach (in commercial applications as well as academic research) is the form of connectionism called *PDP* (parallel distributed processing). *PDP* representations are implemented not by a single symbol in computer memory but by the overall pattern of activity of a network of units. *PDP* processing isn't like classical computation: do this, then do that. Rather, it involves self-equilibrating changes of network activation—and, in learning systems, weight changes on the connections between units.

PDP systems lack the three advantages just mentioned but offer others in compensation. They "naturally" provide content-addressable memory, in which an input pattern automatically reactivates the relevant activity array across the network (as opposed to finding some specific memory address). They allow acceptable pattern-matching performance even if the input pattern is partly missing or accompanied by irrelevant input. And they enable learning by example, as opposed to

learning by being explicitly programmed. All these useful capacities are very difficult to program using classical AI methods.

Neural networks in general (*PDP* and non-*PDP*) are less biologically plausible than is sometimes claimed. Most connectionism bears only a very sketchy likeness to processing in the brain. Nevertheless, some *PDP* models resemble (for instance) specific types of dyslexia, and some non-*PDP* models are based on specific facts of neuroanatomy and neurophysiology.

Most AI researchers today use only classical, or only connectionist, models. Because these have complementary strengths and weaknesses, there is growing interest in "hybrid" models, which try to get the best of both worlds.

Hybrids come in various forms. A connectionist system may mimic, to some extent, the properties of classical AI machines. For example, *recurrent networks*, in which information is fed back from higher to lower levels, can capture (tacitly, not explicitly) some features of hierarchical structure. Conversely, some classical work incorporates bottom-up, parallel processing of subcognitive microfeatures, and the widely used "blackboard" architecture is a sequential-parallel hybrid. And classical and connectionist virtual machines may be combined at a specially designed interface: each part performs the tasks to which it is best suited, control passing from one to the other as appropriate during problem solving.

Both classical and connectionist AI see internal representations as crucial. Recent AI work in situated robotics, largely inspired by insect biology, does not. Classical robots, during planning, must try to anticipate a host of intended consequences and unintended side effects. In general, they can't anticipate everything, so come to grief during plan execution. Even if they are capable of replanning, they have to sit and "think" for a while before moving off again. Meanwhile, anything could happen. But situated robots don't manipulate internal representations of the world: they deal directly with it.

These robots show simple, hard-wired, behaviors triggered by specific environmental cues. One behavior can be inhibited by another (so walking may be inhibited by turning), but that, too, is directly triggered (by hitting an obstacle, perhaps). The robot is

... the unthinkable is now being thought: reconciliation between Anglo-American and Continental philosophy is being considered, partly because of recent advances in AI.

Most AI researchers today use only classical, or only connectionist, models. Because these have complementary strengths and weaknesses, there is growing interest in “hybrid” models, which try to get the best of both worlds.

better adapted to its environment with the higher levels but can function without them. Apparently goal-directed behavior can emerge in this way, without any goal representation or top-down planning. Whether this approach could explain high-level human thinking is another matter.

Evolutionary robotics aims at a still closer biological parallel. Its hardware isn't hand designed but automatically evolved. This approach uses “genetic algorithms” (GAs), widely used in AI for problem solving of many kinds. GAs produce random mutations, or crossovers, in a program's rules. The most useful of the resulting rules, given the task environment, are used (with high probability) for further “breeding.” After many generations, the system may be highly efficient.

For instance, the “brains” of simple robots, and their sensorimotor anatomy, have been evolved in this way. This is an example of work in artificial life (“A-Life”). A-Life studies self-organizing, self-replicating, adaptive systems and (more generally) the emergence of ordered complexity from simple rules. It's closely related to AI. Indeed, because intelligence is a property of living systems, AI might be seen as a subarea of A-Life.

Some Philosophical Controversies

Part of what it means to say that something is philosophically interesting is that it is highly controversial—and AI is.

Many AI theorists believe that certain computational processes are necessary and sufficient for intelligence. John Searle calls this belief “strong AI” and argues that it's fundamentally mistaken. In his view, symbolic AI deals only with syntax, not semantics, so it can't account for intentionality, or meaning. Connectionism can't explain intentionality, either—though it may suggest how meanings are interrelated, once we have them. Besides, Searle claims, it is intuitively obvious that neuroprotein can support intentionality, whereas metal and silicon can't.

Other philosophers are less dismissive. They argue that AI, in one form or another, can help us understand how meaning, concepts, purpose, creativity, and even consciousness are possible. If they are right, then AI (with neuroscience) holds the key to the puzzle of how mechanism can support meaning. If Searle is right, this seductive promise can't be honored.

Contra Searle, we've no particular reason to believe that only neuroprotein can sup-

port intentionality. Moreover, the fact that it does so is highly counterintuitive. Insofar as we understand this at all, we appeal to the information-processing properties of the brain (the role of the sodium pump in enabling message passing, for example), not to its specific material stuff.

As for his claim that syntax alone cannot give us semantics, I agree. Meaning, purpose, and understanding require certain sorts of causal relationship between a system's internal processing and its environment, plus a historical grounding in evolution. Intentionality can be ascribed to an artifact only in a secondary sense.

Causal-evolutionary accounts of intentionality (such as this) aren't universally accepted, however. The concept of “information” is contested, too, so there's no agreement on just what an “information-processing system” is. Further, some philosophers oppose any naturalistic (scientific) account of meaning.

Another philosophical dispute concerns the nature of our mental architecture. Some see thoughts as computations over representations, involving elementary units of the “Language of Thought.” In this view, classical AI is philosophically required, and connectionism is philosophically irrelevant (a mere implementation detail). Other philosophers favor connectionism, as explaining how objective concepts can arise from preconceptual thought, how understanding rests on prototypes and family resemblances, and how scientific explanation is possible. Yet others refuse to join the fray, seeing no fundamental philosophical difference between these two AI approaches.

What of consciousness? Suppose AI were to produce a robot outwardly just like us, passing the Turing Test in whatever form we cared to pose it. Would such a robot be conscious?

Searle believes that a robot made of inorganic materials couldn't be conscious. But at least he allows that neuroscience might explain consciousness. Some philosophers despair of any scientific explanation of it at all. They argue that subjective consciousness (what it's like to be an experiencing subject) can never be captured by objective descriptions. Or they argue that the mind-body problem is forever beyond our capacities to solve, much as quantum physics is beyond

the understanding of a dog. Comparatively, Searle is an optimist—about neuroscience, if not about AI.

The greatest optimist about AI consciousness is Dan Dennett. He explains the richness of our experience in terms of highly sensitive behavioral discriminations and idiosyncratic associations. Phenomenal experiences “as such” simply don’t exist. Or rather, because he admits he can’t strictly prove this, there’s no good reason to believe that they do. He even believes that a conscious robot may be built in the near future. (Its consciousness would be unlike ours, and much less complex, but the same applies to nonhuman animals.) This belief strikes many people as absurd—but that’s not to say that anyone else has offered a convincing philosophy of consciousness.

As though all these philosophical disputes weren’t enough, neo-Heideggerian murmurings are afoot. They threaten the fundamental assumptions of AI, for they reject the subject-object distinction presupposed by materialists and idealists alike, and deny the epistemological primacy of science.

Heideggerian critiques of AI aren’t new. But they’re now being mounted by people sympathetic to computer modeling: in particular, to situated and evolutionary robotics and to A-Life’s study of “animats.” These people see organisms as dynamic systems closely coupled with their environment. Instead of positing internal representations of an objective external world, they speak of whole systems embedded in, and adapted to, their own particular “worlds.”

The Future of AI

Our friendly astrologer would be hard put to forecast the future of AI, even if a precise time of birth were agreed.

The AI of the 1990s is intriguingly diverse. Various scientific bets are now being laid, and it’s not obvious (though it may be true) that some eclectic hybrid will eventually enable all to win. Moreover, new AI concepts will undoubtedly emerge, as research develops. And the philosophical difficulties are legion. It’s unclear, for instance, how neo-Heideggerian critiques will affect scientific AI, and whether connectionist and/or situated AI are acceptable ways of grounding meaning in naturalistic terms.

With respect to final verdicts, then, both juries—scientific and philosophical—are still out.

Nevertheless, AI has given direction (and

clarity) to many psychological projects. As McCulloch and Pitts predicted, it has revived the philosophy of mind. And the unthinkable is now being thought: reconciliation between Anglo-American and Continental philosophy is being considered, partly because of recent advances in AI. Not bad, for the first half-century.

Note

1. This article is based on a longer paper written for a joint discussion meeting of the Royal Society and British Academy, on the theme “Artificial Intelligence and the Mind: New Breakthroughs or Dead-Ends?” The papers from the meeting, edited by M. A. Boden, A. Bundy, and R. M. Needham, were published as a special number of the *Philosophical Transactions of the Royal Society, London* (Series A, Physics and Engineering) 349:1–166.



Margaret A. Boden is professor of philosophy and psychology at the School of Cognitive and Computing Sciences, University of Sussex, England. She is a fellow of the British Academy, a member of the New York Academy of Sciences, and a fellow of the American Association for

Artificial Intelligence. Her books include *Artificial Intelligence and Natural Man* (second edition, 1987, Basic Books), *The Creative Mind: Myths and Mechanisms* (1991, Basic Books), and *The Philosophy of Artificial Intelligence* (1990, Oxford University Press). Her edited volume, *The Philosophy of Artificial Life*, will be published in 1996 by Oxford University Press.



If You're Serious about AI,
You Should Be a Member of

AAAI

the Premier AI Society

445 Burgess Drive • Menlo Park, CA 94025
(415) 328-3123 • (415) 321-4457 (fax)
info@aaai.org