# Book review

## Action editor: Stephen Wermter

*Artificial Minds*, Stan Franklin, MIT Press, Cambridge, MA, 1995, xi + 449 pp., ISBN 0-262-06178-3 (HB), 0-262-56109-3 (PB)

Stan Franklin's book *Artificial Minds* explores different ways of understanding minds. It surveys philosophies of mind, signs of mind in the animal kingdom, the use of computers for modeling life and mind's functions, brain activity that could underly mind, robots as embodiments of minds, and the debate and controversy surrounding different points of view; and it considers the possibility of building systems with artificial minds. The book is appropriate for an educated reader with an interest in such matters as well as the professional, thanks to its style, scope, and depth. After you have read it once and absorbed the material, read it again for the insight that comes from seeing the work as a whole. It is well worth it.

Our popular notion has it that a mind is something particularly human, exemplified by conscious thinking, feeling, emotions, reasoning, intuition, remembering, communicating. Things relating to the mind are called mental and are contrasted with the physical; mind being contrasted with matter. Against this picture the idea of artificial minds may seem radical, even heretical. Not only would bears and birds and butterflies have minds, but we could build machines with minds, which is more controversial than machines with intelligence.

Perhaps artificial intelligence is more acceptable to us than artificial minds because we have more of a handle on intelligence — we measure it crudely with IQ tests, for example — and because the field of Artificial Intelligence, or AI, has been around longer. For more than 30 years AI has been taught in leading universities giving it legitimacy even if most of us do not know how AI works. It is a little like our accepting computers without knowing how they work. How many of us could build a computer from delay lines and logic gates, from relays and switches? Is it even necessary to know computers at that level of detail, unless you are an engineer? I believe that it is doubly so, if we are to understand artificial minds: first because the computer is at the core of many models of mind, as shown in Franklin's book, and then to drive home the idea that when we are looking for an underlying mechanism of X, we are looking for something very unlike X; computers allow us to see and to reflect on complex and obscure behavior arising from mechanisms that we fully understand. Such relating of behavior to mechanism is central to Franklin's *Artificial Minds*.

The study of the mind has historically been the domain of philosophy and more recently of the *humanities*, which Webster (10th Collegiate) defines as ''the branches of learning (as philosophy, languages) that investigate human constructs and concerns as opposed to natural processes (as physics or chemistry).'' Again we see the mental contrasted with the physical: mind as not resulting from natural processes. However, early Western philosophy included also physical and biological science — Aristotle felt qualified to do it all — that is, before the mechanisms of astronomy, physics, chemistry, metabolism, and heredity, in roughly that order, were known anything like how we know them now.

Western thought is marked by a passage from ignorance through philosophy into natural sciences, as more and more complex phenomena, in roughly the above order, have been explained in terms of underlying mechanisms. Franklin's book follows this

trend by taking us to the next frontier: to explaining minds in terms of underlying mechanisms. Undoubtedly there are frontiers beyond this one, such as those studied in the social sciences and frontiers yet unknown, but this one is special. We have arrived at a point where the thing being explained is just as complex as the thing doing the explaining; where the mind is trying to sort out its own mechanism.

Is this asking for the impossible, like lifting ourselves by our bootstraps? Although this is fertile ground for philosophical debate, and Franklin reviews some of it for us, the issues are not resolved by debate and thought experiments alone. We must also gather facts and evidence on how minds work and fail. But even that is tricky when we ourselves are so completely immersed in the phenomenon we try to understand. Just think of the difficulty that humankind has had in seeing ourselves as minute specks on a tiny planet rotating around its axis and whirling around the sun in a huge and indifferent universe, when we experience ourselves as being at the center of it all, on stationary Earth that provides for our needs, with the heavens revolving around us. The illusion is irresistible — because it is our reality. It is even more so with minds. Each person's mental life is that person's reality, and what gives rise to it — what underlies mental life — is hidden from a person's mind.

*Third-person view*. How does Franklin navigate the ''illusions'' arising from deep self-involvement? Mainly by insisting that mind is not uniquely human but a product of evolution, so that the animal kingdom offers us countless varieties and degrees of mind. Simple animals have simple minds, animals most like us have minds most like ours, and an infant's simple mind develops into an adult's more complex mind. This gives us a third-person view of mind and turns the basic question about minds into: When should we regard something as having a mind? Rather than trying to define it sharply by necessary and sufficient conditions, we take the human mind as a standard and humanlike behavior as an indicator of mind: the more like us something acts, the more of a mind it has. This also is the logic behind the test that Turing has proposed for deciding whether a machine can think.

In Franklin's (third-person) view, the working of a mind is seen in action or, more precisely, in the *interaction* of a thing with its environment. The question then becomes: How is it accomplished? How does sensing lead to action? What internal structures and functions are indicated? The underlying premise is that organized activity of some kind within the individual is necessary. With humans and animals we take it to be the activity of the brain. This approach leads into examining physical structure — the organization of matter — as a prerequisite of mind and, finally, into the possibility of artificial minds.

*Engineering view*. Franklin classifies the study of mind according to two criteria, top–down vs. bottom–up, and analytic vs. synthetic. He sympathizes most strongly with the bottom–up, synthetic approach, which he calls ''Mechanisms of Mind ''; the book is indeed an exploration of the mechanisms of mind. According to it, we take simple components whose working we understand and from them design a system that works so like a mind that it surprises even its designer — in other words, it has lifelike emergent behavior. On this, Franklin quotes Carver Mead: ''If we really understand a system we will be able to build it. Conversely, we can be sure that we do not fully understand a system until we have synthesized and demonstrated a working model. '' I call this the *engineering view* and will expand on it.

Engineering does not rank highly in intellectual debate — nowhere near philosophy or art or mathematics or the nature of consciousness. Like agriculture, it seems utilitarian, mundane, intellectually uninteresting. Why bother with neurons and synapses, or with delay lines and logic gates, when we are talking of minds? And that exactly is the reason! Minds are exceedingly difficult for us to grasp because their underpinnings are invisible to us, and so we talk about minds in high-level, abstract terms such as beliefs, desires, intentions, motivation, plans, and goals. This gives us a feeling that we really understand minds, but when we try in AI to design minds from such abstract building blocks, we are not particularly successful. According to the engineering view, we are far from understanding minds. Thus, the engineering view is an exacting standard against which to judge the depth of our understanding. It disciplines our minds and fosters

intellectual honesty and, as things stand, it gives us cause to be humble about the depth of our understanding.

In yet another way engineering can lead to understanding minds. Brains are too complex — there are too many details — for us to sort out, and duplicating a brain does not equal understanding it. However, the principles by which brains accomplish their feat need not be many and complicated, they just are effective and non-obvious. Once the *principles* are understood, we can build systems based on them and thereby learn how minds work.

*Autonomous-system view*. Franklin views minds as control systems of autonomous agents. They allow agents to learn, and so we need a plausible account of learning. Learning is sometimes likened to programming, with the brain corresponding to computer hardware and the mind to software. However, many of our programs for mindlike functions are based on our abstract understanding of minds. In doing so, we build *our* meanings into the system and thereby become a part of the system and lose a third-person view of it. By definition, an autonomous agent has its *own* meanings — Franklin expresses it by saying that minds create information — and anything resembling a program comes into existence through the system's own actions and interactions with its environment. This agrees with the Enactive Paradigm of mind put forth by Varela, Thompson, and Rosch and reviewed sympathetically by Franklin. From raising children we know that explicit instruction, which is the mind's equivalent of being programmed from without, is inefficient compared to learning from example. It shows that natural minds learn very differently from today's programmed computers, so that equating the (human) mind with software is more like a cliché than a helpful analogy. I doubt that minds can be truly appreciated and understood without understanding also the hardware.

The topic of *meaning* brings us to looking at symbolic representation through the eyes of biologists, engineers, and roboticists. Human language is a paradigm of symbolic representation: words are arbitrary patterns for meanings, and the meanings are something else somewhere else. Contrast this with the following. An autonomous system is coupled to its environment by sensors and effectors; in an animal it is by huge arrays of neurons, and everything about the world that is available to the animal's mind is defined by neural firing patterns. We can think of these patterns as the animal's symbols, except that they are not arbitrary. Some patterns make the animal move, some are pleasurable, some painful, and so forth, and we can no longer distinguish symbol from meaning. Similarly with artificial autonomous agents, some patterns over some of the gates and delay lines are the agent's own meanings: there is no representation in the traditional AI sense. This idea is taken to its extreme in Brooks' Nouvelle AI and Subsumption Architecture, in which Franklin detects the seeds of the Third AI Debate.

The book describes all three. The first AI debate was over the possibility that a machine, a computer, could think. The second pitted connectionist AI against traditional symbolic AI, the connectionists rallying over the idea that artificial neural nets are more brainlike than computers, and the others arguing that it did not matter. The third AI debate is about the necessity and nature of mental representations. The old and the nouvelle AI would have us in opposite corners, when the interesting action is more likely to be all over the playing field. Franklin hints at the possibility of varieties and degrees of symbolism, akin to varieties and degrees of mind. I agree with that view so much so that it defines my present research. It seems to me that arbitrary symbols are not needed for the things we do as infants to satisfy our basic needs, whereas our making up, telling, and understanding of stories, for example, implies internal modeling of the world and sophisticated use of more or less arbitrary symbols. The internal modeling could be partly traditional symbolic AI-ish but it cannot be fully so, and so we must look for representations that allow for both arbitrary and meaningful symbols, and everything in between. Pollack's RAAM and Chalmers' experimenting with it, which Franklin reviews, are steps in that direction.

Through the book runs a philosophical undercurrent that leads to Franklin's Action Selection Paradigm of mind, which is summarized in the following three statements (p. 419): (1) Cognition is the process by which an autonomous agent selects actions; (2) actions emerge from the interaction of multiple, diverse, relatively independent modules;

and (3) a cognitive system functions adequately when it successfully satisfies its needs within its environment. That is not meant as a recipe for engineering but as criteria by which to gauge philosophies and models of mind. They are to be judged from the viewpoint of an autonomous system that interacts with its environment, and the success criterion means that things with minds could come to be by evolution.

In addition to controlling action, the human mind is occupied with all sorts of things for their own sake: we sing and dance and chat with friends and produce plays and go to the theater and watch the sunset for the pleasure, stimulation, and peace of mind that they give us. There must be more to minds than what meets the third person's eye. My reading of *Artificial Minds* suggests that Franklin would agree.

Franklin's notion of mind does not necessarily culminate in the human mind. Some people — Franklin cites Moravec, and Margulis and Sagan — speculate that once mind's mechanisms are understood we will build silicon minds more powerful than ours. We do not even have to go beyond what already exists. A family, a community, a corporation, a nation, and all of humanity can be thought of as autonomous agents with superminds that are harder yet for us to comprehend than an individual human mind. Such superminds pose an interesting puzzle: When the idea of mind as a society of agents is taken to its logical conclusion, it obscures rather than clarifies the notion of mind and mind's mechanisms.

Franklin's writing is admirable. It is thoughtful, lively, informative, and accessible. He explains the working of more than a dozen models. I already knew some of them well — one being my own — and Franklin's descriptions of them gives me faith in the quality of instruction I received on the others. He shows concern for the readers' difficulties with unfamiliar subject matter, and so he provides background material, is generous with diagrams, figures, and tables, avoids specialized language, and steers around ambiguity with examples. It is a precious gift to the reader not to have to struggle with the presentation when struggling with new concepts and points of view is challenging enough. I noticed only a few instances where his watchfulness gave way to his long and intimate familiarity with the subject

matter. To characterize the mind (or language or consciousness or free will) as ''scalar '' or not ''Boolean'', or to reject the notion of a ''*linear continuum*'' of minds (p. 412) is mathematicians' shorthand that can distract others.

The book is very helpful when you want to learn more. Numerous comments throughout the text, together with 14 pages of references, give a wealth of pointers to further reading. To the reading list I would like to *add Brains, Behavior, and Robotics* by James Albus (Peterborough, NH: Byte Books, 1981), *Neurons and Symbols: The Stuff that Mind Is Made of* by Igor Aleksander and Helen Morton (London: Chapman & Hall, 1993), and Aleksander's *Impossible Minds: My Neurons, My Consciousness* (London: Imperial College Press, 1996). All three take an engineering view and an autonomous-system view similar to Franklin's.

The production of a book is an endless battle against trivial errors, and we rarely claim total victory. There are a few wrong words that spelling programs did not catch; in figures 11.3–11.5 the arrows for successor and predecessor links have been interchanged; some page numbers in the index are off by one; and 20 bifurcations would give Moravec's bush robot (Fig. 15.3) a mere million cilia rather than a trillion — did he mean 20 ''tetrafurcations''? The reader can easily disregard such errors.

I will conclude with a personal view on the state of our science as evident in the observations, experiments, and models that Franklin writes about. Our models, both symbolic and connectionist, are abstract at too high a level. They are more like metaphors for mind than basic mechanisms, they act more like mirrors than microscopes. They are generalizations from what we see and introspect of mind's working. The problem with such abstraction is that even when it credibly describes behavior, it does not constrain sufficiently the underlying mechanisms. It does not tell how brains and their models should be constructed, and so we get very little in terms of emergent properties from our models, very few pleasant surprises. Our models hardly begin to explain how brains learn and minds develop.

Since the underlying mechanisms are hidden they have to be inferred and then tested by modeling, and so the power of our minds to imagine and to

conjecture is crucial. That power comes from experience. We need to know what psychology and neuroscience can tell us about the mind's working, and we need to know mathematical systems. For example, a mathematical model of a certain kind will be suggested and understood only by someone who is familiar with that kind of mathematics.

But does our understanding of computers, or minds, have to go down to the level of delay lines and logic gates, or neurons and synapses? In the following sense it does. They make it plain how unlike the infrastructure is from the phenomenon we want to understand, being it the behavior of a computer or a mind. These components and their actions are meaningless in the sense that the whole system behaves meaningfully. So the *organization* of the components is crucial and is what we must understand. Notice also that delay lines and logic gates refer to principles of operation — namely, to holding and combining of data, holding and combining of patterns — except that these principles are not mysterious to us the way minds are, and they are easily realized in physical devices. I see a close analogy to understanding chemistry in terms of atomic structure, or life forms in terms of chromosomes, genes, and the genetic code, and I bet that we will never give them up for the old ways of thinking about matter or heredity.

Constructive modeling starts with simple com-

ponents, which are then built into circuits for working with patterns. The patterns realize abstract states, and the circuits govern state transitions, accounting for the system's behavior. Some patterns or states are meaningful through grounding, for which Brook's subsumption architecture is a model, and others become meaningful by composition, which in some ways must resemble symbolic processing. All of this, too, is abstract but it is abstract at a low level and the high-level abstractions — new concepts — are based on the low-level ones. The notion that implementation does not matter is wishful thinking for a scientist.

Major discoveries and hard work lie ahead before we uncover a foundation for the working of the mind that is anything like chromosomes, genes, and the genetic code are for the working of heredity and the evolution of life. Yet we must try, or accept the alternative that minds work by magic. *Artificial Minds* is an excellent introduction to the ways of our trying.

Pentti Kanerva
*RWCP Theoretical Foundation SICS Laboratory*
*Real World Computing Partnership*
*Swedish Institute of Computer Science*
*Box 1263, SE-164 29 Kista*
*Sweden*