Artificial Minds

Stan Franklin

Artificial Minds

Stan Franklin

A Bradford Book
The MIT Press
Cambridge, Massachusetts
London, England

First MIT Press paperback edition, 1997

This book was set in Sabon by Graphic Composition, Inc., Athens, Georgia and was
printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Franklin, Stan.
Artificial minds / Stan Franklin.
p. cm.
"A Bradford book."
Includes bibliographical references and index.
ISBN 0-262-06178-3 (HB), 0-262-56109-3(PB)
1. Artificial intelligence. 2. Cognitive science. 3. Brain. I. Title.
Q335.F733 1995
006.3—dc20 94-38796
CIP

Dedicated to the memory of Lil Franklin and of Paul Byrne

Preface

For thousands of years humans have struggled with the relationship between the mental and the physical, with the mind–body problem. Our efforts have been rewarded with a clear understanding of some of the alternatives and a great respect for the difficulty of the problem, but with no satisfactory solution. At long last, pieces of a solution seem to be on the horizon. Researchers in many fields are producing possible mechanisms of mind. Some are theoretical, some algorithmic, some robotic. Together they are leading us to a very different way of thinking about mind, a new paradigm of mind. This work is an exploration of these mechanisms of mind, and of the new paradigm of mind that accompanies them. It's intended to change your mind about mind.

Though the research involved is highly technical, this exploration of it is not. Technical concepts have been purposefully made accessible to the lay reader. Though I hope and expect that artificial intelligence researchers, computer scientists, cognitive psychologists, ethologists, linguists, neurophysiologists, philosophers of mind, roboticists, and such can learn of activity in the other fields from this multidisciplinary tour, it's not designed only for them. I also hope that anyone interested in how mind arises from brains, or from machines, can enjoy accompanying us on this exploration. Finally, I hope to have achieved accessibility without damage to the many powerful and elegant ideas, concepts, models, and theories to be explored.

Preparations for this exploration of mechanisms of mind began with a series of lectures titled "Artificial Minds." Even before the lectures began, it was becoming clear that my role was to be as much that of hawkster as of tour guide. I had my own goods to sell to the audience, an emerging new way of thinking about mind. The tenets of this new paradigm helped guide the selection of the mechanisms to be explored. Each mechanism is to provide background for understanding the new paradigm, or evidence to support it. I wanted to report on these mechanisms and to have a hand in shaping this new paradigm.

Little did I realize the magnitude of this endeavor. It's now four years later, and the end is in sight only with the help of more people than I would have believed. Appreciations are in order.

Thanks to Don Franceschetti for suggesting the title "Artificial Minds."

My heartfelt thanks to the participants in the Cognitive Science Seminar during the summer and fall of 1991 for listening to my "Artificial Minds" lectures and arguing with me about them. Thanks also to the students in the "Artificial Minds" class in the spring of 1994 who listened to and commented on these same lectures, and a few more.

The University of Memphis, through its College of Arts and Sciences and its Department of Mathematical Sciences, has been supportive throughout, by providing a sabbatical during the fall of 1991 and a faculty research grant during the summer of 1994.

Thanks to Elizabeth Bainbridge, to a second typist I never met, and to Helen Wheeler for transcribing the recorded lectures. Thanks also to the Institute for Intelligence Systems and to the Mathematical Sciences Department for supporting this work. Also thanks to Jarek Wilkiewicz for helping me to obtain permissions to reprint, for proofreading, and for indexing, all supported by the Institute.

For invaluable conversations over many years about this material, my thanks to Kaveh Safa, Daniel Chan, and Art Graesser.

And many, many thanks for stimulating conversations over the years to participants in the weekly AI lunch: Bill Baggett, Paul Byrne, Max Garzon, Art Graesser, David Kilman, Jim Michie, Joel Neely, Lloyd Partridge, Bob Schreiber, Shane Swamer, and many others. So often I tried ideas out on you first. And so often they came away improved, or discarded.

Perhaps my deepest indebtedness is to those friends (and relatives) who read and commented on chapters as they were produced: Bob Sweeney, Art Graesser, Phil Franklin, Dan Jones, Bill Boyd, David Kilman, Pat Patterson, David Lee Larom, John Caulfield, Nick Herbert, and Elena Franklin. Though I revised assiduously in the light of their comments, as the

book goes to press, I find myself wishing I'd heeded even more of what they said. That would surely have improved the work, but perhaps would have required a second volume.

And thanks also to several reviewers—Mark Bedau, John Caulfield, Janet Halperin, John Holland, Chris Langton, and others who remain anonymous—for their helpful comments and suggestions.

My appreciation to Larry McPherson and to Eric Ehrhart for encouragement and badgering, to Brian Rotman for encouragement and for discussions about writing and publishing, and to Ralph Randolph for advice about how to look for a publisher.

Thanks to Fiona Stevens and to Harry and Betty Stanton for encouraging this project, and getting it through the door at the MIT Press. Additional gratitude to Fiona for so patiently dealing with my many questions and concerns, and for prodding me actually to finish (stop) the manuscript. I'm sure the phone company appreciates her also. And my appreciations to Katherine Arnoldi for gently guiding me through the copyediting process, and to Beth Wilson for a superb job of copyediting.

Much appreciation to Arthur Goodman for such fine work on the illustrations.

Finally, and most important, I'm so grateful to my wife, Jeannie Stonebrook, for support and encouragement, for sage advice, and for assuming many of my chores to give me time for this project.

Contents

1

Mechanisms of Mind

Phase Transitions and Fascinating Questions

The title *Artificial Minds* has a strange ring to it, almost oxymoronic. Are there any artificial minds? Could there possibly be artificial minds? What would make you think there might be such things to begin with? Why would one want to ask such a question? Let's take a brief detour by way of putting these questions in broader perspective.

Recall the notion of phase transition from that almost forgotten physics course. Common phase transitions occur when ice melts to water, shifting from solid to liquid, and when water boils to steam, shifting from liquid to gas. The properties of systems change quite rapidly at these phase boundaries. Some would say that all the interesting stuff, including life itself, happens only at phase boundaries (Langton 1992a). But that's another story.

Three questions seem to be inherently interesting, even fascinating, to many people:

1. How did the universe come to be?
2. How did life originate?
3. What is the nature of intelligence?

Each of these apparently has to do with a phase transition. For the cosmological question, there's the phase transition between being and not being, or between pre- and post-big bang, if you're of that persuasion. The origin of life question focuses on the phase transition between living and nonliving matter. Asking about the nature of mind leads to the phase transition between the physical and the mental.

These phase transitions are not as sharp as those bounding solids and liquids, or liquids and gasses. Some might argue that no phase transition occurs between being and not being, because the class of nonbeing has nothing in it. Also, the dividing line between the living and the nonliving isn't easily agreed upon. Biologists might argue about whether viruses are alive. How about the boundary between the physical and the mental? That question is the major issue of this book.

Here we will be primarily concerned with mechanisms of mind, with how mental activity arises from physical substructure. Don't miss the underlying assumption of the last sentence. It takes a currently fashionable position on the mind–body problem. More on that a little later, and in chapter 2. For now, I hope that our detour into the natural questions arising from phase transitions has begun the booklong process of putting the nature of intelligence in context.

Life Itself

Mind, until now, has been associated with life, usually only with human life. If we're to explore the mechanisms of mind, it would be well to trace a little of their history and development. Focusing a wide-angle lens on life, as we know it on Earth, may help. This goal gives me a perfect opportunity to tout you onto a marvelous little book by the DNA decoder, Francis Crick, titled *Life Itself* (Crick 1981). A wag of a reviewer referred to it as "Life Itself by Crick Himself."

Figure 1.1, taken from *Life Itself,* displays a time line of the universe. The Age of Humans is not explicitly mentioned. Why not? Well, from this view we're hardly visible. If the life span of the Earth to date, roughly 4.5 billion years, was represented by the Eiffel Tower, our "dominion" would occupy the thickness of the uppermost coat of paint. Compare this with what Crick calls the Age of the Prokaryotes, the simplest kind of single-cell organism, including bacteria and blue-green algae, which have had the Earth to themselves for well over half its lifetime. Some would say that neither humans nor ants nor cockroaches dominate the Earth today. It's the Age of Bacteria. It always has been, and always will be, as long as there's an Earth (Margulis and Sagan 1986).

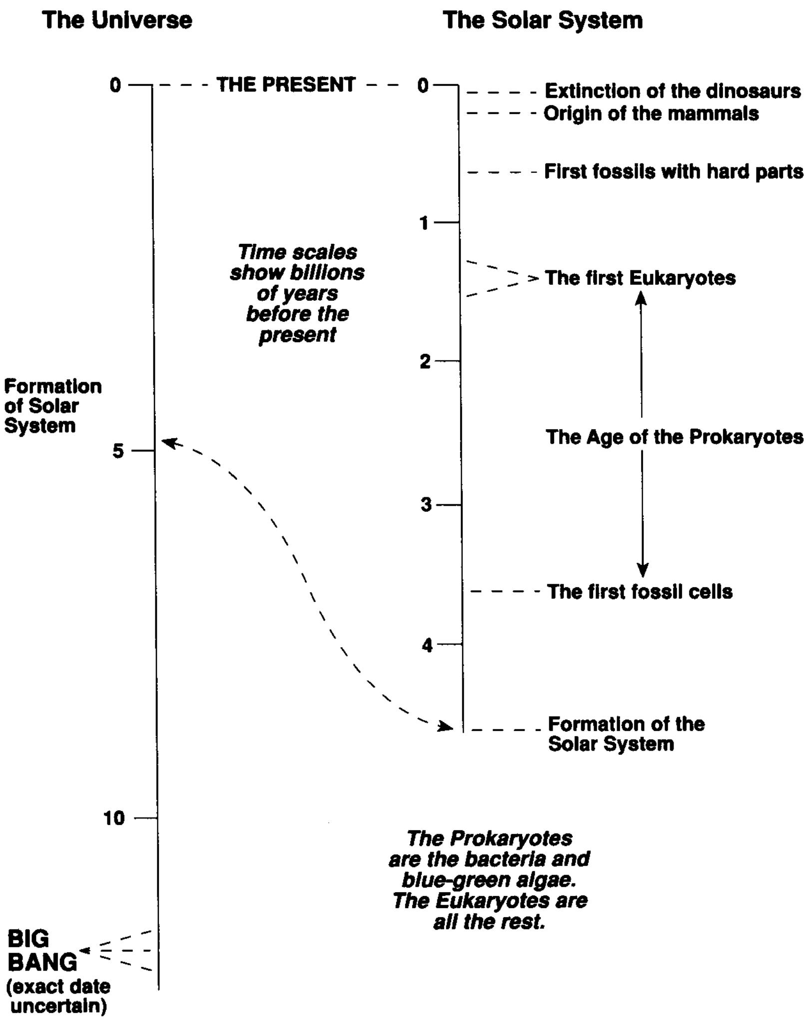

Figure 1.1
Life in time (redrawn from Crick 1981)

Bacteria are small, one to a few microns in diameter. They can sense food concentrations and toxic chemicals, and move toward or away from them by means of tiny flagella. They exercise efficient control of their metabolism, and enjoy a tenuous sex life without being dependent on sex for reproduction. Bacteria thrive in every ecological niche on Earth, inhabiting even nuclear reactors. Their combined biomass probably exceeds that of all other living things. Loose talk of humans wiping out life on Earth by nuclear warfare is just that: loose talk. Bacteria will survive quite nicely, thank you.

Amid all this glorifying of the infinitesimal, let's ask if in any sense bacteria could be said to partake of mind. (This is not the most ridiculous question we'll pose, but it's close.) The knee-jerk answer of "Certainly not!" may be less satisfying than you would think, after we've finished our exploration.

Next up on Crick's chart we find the first eukaryotes. These include all cells containing a well-defined nucleus and organelles surrounded by a membrane. Some eukaryotes are single-celled, like amoebas. All multicell organisms, like you and me and the grass on your front lawn, are colonies of eukaryotic cells.

At long last we arrive at the stuff of paleontology, the first fossils with hard parts. Our brief tour of the Earth's time line à la Crick ends with the origins of mammals and the extinction of the dinosaurs. (Don't tell the birds, who are widely believed to be the descendants of the dinosaurs.)

What strikes me most in Crick's chart is that mind, as we're accustomed to think about it, evolved from bacteria very much as a latecomer, even as an afterthought. Mind, in this view, has had no affect on life except during this last coat of paint atop the Eiffel Tower. Some trivial chance fluctuation along the way could easily have resulted in no humans[1] and, hence, no mind according to the *Oxford English Dictionary*.[2]

Crick makes two other points that may well prove germane to our exploration. One has to do with natural selection, and the other with what he calls the "combinational principle." Let's look first at Crick's combinational principle. Here's the argument. All life as we know it is complex. Nothing but natural selection produces such complexity.[3] This complexity requires the storing and replication of much information. The only efficient mechanism is the combinational principle: express the information by using a small number of standard units combined in very many different ways. We use the alphabet in this way to produce written language. Perhaps a more pertinent example is DNA, whose constituents form the "alphabet of life." If one wants artificial minds to replicate, this combinational principle doesn't seem like such a bad idea.

Here's Crick's most interesting view of natural selection: what natural selection does is to make rare chance events become common. Suppose you have a genome with a mutation in it, a rare, improbable event. If the mutation is successful, which would also be rare, it can influence organisms down the eons and eventuallly become quite common. Given enough time, organisms become finely tuned to their environment. This works as long as the environment stays fairly constant, but major environmental changes can lead to extinctions.

Crick points out that there's no mechanism to direct changes in the genes so that favorable alterations are produced. You don't know which mutations are going to be successful and which are not. This point has recently become controversial. There's new evidence that some bacteria (what else?), when placed in a sugar solution, mutate in the direction of metabolizing sugar at greater than chance rates. Some people are beginning to think there might indeed be mechanisms directing evolution. The issue of mechanisms directing evolution provides a bridge to get us back on track from our detour.

Stubbs's Great Leaps

Most evolutionists staunchly deny any direction or progress in evolution, any mechanism providing direction. Nonetheless, Monday morning quarterbacks often discern direction in the evolutionary record. One such is Derek Stubbs, a physician-turned-computer scientist, who publishes a newsletter now titled *Sixth Generation Computing*. Stubbs maintains that the evolution of life has made several "great leaps" when judged by "the criterion of adaptability and particularly the speed of adaptability" (1989).[4] A more precise criterion might be rate of adaptability. Keep in mind that "adaptability" is commonly used in several senses in this one context. You can speak of evolution as adapting species to their environments. You can also think of an individual adapting its behavior during its lifetime via learning. You may also talk of an individual adapting its behavior to short-term changes in its surroundings without new learning. These three meanings of the word are quite distinct. Watch out for the potholes.

So what are the "great leaps"? Stubbs identifies seven.

Life itself is the first great leap. First came the ability to reproduce, the basic definition of life.[5] Mutations are blueprint errors that, along with transcription errors,[6] allow for variation from which to select on the basis of fitness.

Second, sexual reproduction "allowed organisms to swap great chunks of adaptive blueprints from their selected-as-fit mating partners."

Third, multicell organisms developed for "safety in numbers" and for "the enhanced productivity of division-of-labor." (Each of the eukaryotic cells comprising our multicell organism contains organelles that some think derived from independently living prokaryotic ancestors [Margulis and Sagan 1986; Gould 1989]. These were incorporated into the eukaryotic cell. Perhaps they were parasitic to begin with, and evolved some sort of symbiosis. Hence we might view each eukaryotic cell as multicellular.)

The fourth leap was the development of specialized nerve cells that allowed the organism to "find food and mates and escape predators more rapidly."

Fifth was "Invention of a central nervous system. This allowed the organisms to more rapidly adapt to the highly nonlinear dynamics of the outside world and to store memories."

So far these "great leaps" seem uncontroversial, keeping in mind the criterion of rate of adaptability. But hang on to your hat for the next one. Stubbs says:

> Then nothing happened for half a billion years until one creature invented computers . . . and shortly thereafter, the sixth step of evolution was invented as artificial neural networks: artificial learning machines. Life was becoming a going concern and was really in business to start shaping the Universe. Competitors such as rocks and gravity and earth, wind, and fire were taking a distinctly second place.

We'll hear a lot more about artificial neural networks in this and subsequent chapters. For now, suffice it to say that they are computing devices based loosely on biological models. You may or may not agree with

Stubbs's assessment of artificial learning machines as a great leap in evolution. But it's clear that such machines must partake in at least one aspect of what we call mind: learning.

At this point Stubbs leaves the past and leaps into the future: "From here onwards, life was guiding its own evolution, since the seventh step is genetic engineering, environment engineering and life-computer symbiosis." One might infer the coming of artificial minds from "life-computer symbiosis," or one might dismiss the whole notion as science fiction. Some competent scientists are predicting artificial minds. As a mathematician, like me, is wont to do, let's look at one of the extremes.

Mind Children

Among the competent scientists who, after serious study, expect to see artificial minds and then some, is Hans Moravec, a roboticist at Carnegie Mellon University. The following quote from his book *Mind Children* sets forth his expectations clearly:

> Today, our machines are still simple creations, requiring the parental care and hovering attention of any newborn, hardly worthy of the word "intelligent." But within the next century they will mature into entities as complex as ourselves, and eventually into something transcending everything we know—in whom we can take pride when they refer to themselves as our descendants. (1988, p. 1)

Keep in mind that Moravec is a serious scientist expressing a thoughtful judgment backed by a book full of arguments. He may well be wrong, but he's not to be lightly dismissed. Toward the end of this volume, we'll have a look at some of his arguments. In the meantime, let's look at one more quote from Moravec to remove all doubt about exactly what he's predicting.

> We are very near to the time when virtually no essential human function, physical or mental, will lack an artificial counterpart. The embodiment of this convergence of cultural developments will be the intelligent robot, a machine that can think and act as a human. (1988, p. 2)

We now have an existence proof. There is at least one serious scientist who thinks it's reasonable to talk of artificial minds. There are others.[7] I hope there will be many more after this book has been widely read. (Moravec shouldn't be allowed a monopoly on wild predictions.)

Moravec has also engaged in Monday morning quarterbacking, discerning a direction through prehistory leading (inevitably?) to mind children (1988, p. 2). A graphical version appears as figure 1.2. Notice how what each of us sees depends so heavily on our interest and concerns—that is, on the ax we're grinding. Figures 1.1 and 1.2 overlap almost not at all. Such time lines, like statistics, can be bent to almost any will.

What Is Mind, and Why Study It?

We started out asking if there were any artificial minds, or if there even could be. So far, our best answer is that some respectable scientists believe there can be. That's not very convincing. Some respectable scientists can

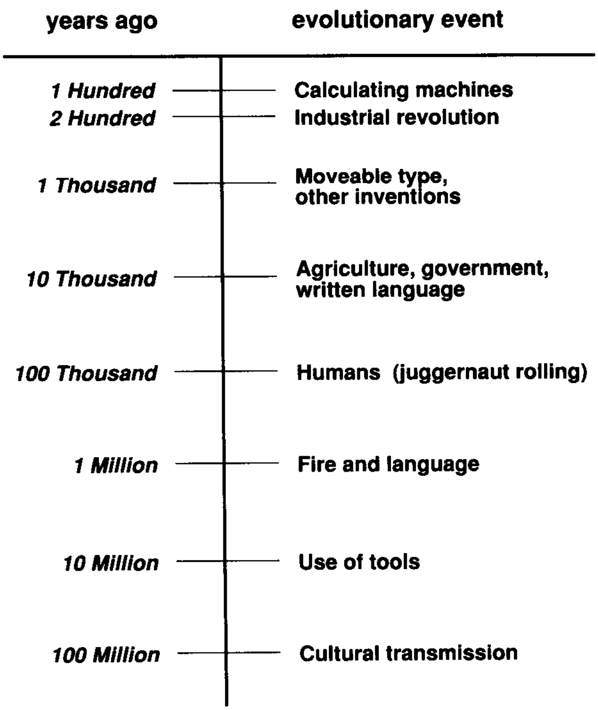

Figure 1.2
Evolution of mind children

be found who believe almost anything. I hope we can do better before the end of our exploration.

But even if we could be more convincing about the possibility of artificial minds, there's still the question of why we should study them. Why explore the mechanisms of mind? The preceding sections have provided answers, at least one of which, the first, is convincing to me. The three reasons are:

1. Questions of the nature of intelligence are inherently fascinating. The study of artificial minds may well throw light on this question.
2. Stubbs might well suggest that we study artificial minds, at least artificial learning machines, to better understand the coming man-machine symbiosis.
3. If we give credence to Moravec's predictions, we would be well advised to study artificial minds to prepare to give birth to our upcoming mind children, and to deal with them more effectively.

Now suppose we are at least marginally convinced by these arguments and, therefore, bent on exploring mechanisms of mind. What does this mean? What is mind? How can mental events occur in a physical world? Do mental events arise from the physical, or are they some kind of spiritual or mental stuff of their own? If they do arise from the physical, how do they do it? And if they don't, where do they come from? Later in this chapter we'll take a brief excursion into the philosophy of mind and the mind–body problem. Chapter 2 is devoted to these issues. For now, let's make what is known as the physicalist assumption: *Mind is what brain does, or something very like it in relevant ways.*

I want to assume so, not because it's the only reasonable alternative but because it allows a scientific approach to the mind–body problem.[8] Let's act on the physicalist assumption and see how far we can get. Let's ask how mental events arise from the physical, and what the mechanisms of mind are. Can they be simulated, or even implemented?

Having focused on these questions, how can we hope to answer them? There are approaches to the study of mind from various disciplines. Figure 1.3 displays some of them via their top-down versus bottom-up and their synthetic versus analytic dimensions.

You can start from *cognitive psychology,* which is a *top-down* approach. By and large it studies high-level concepts. It also takes an *ana-*