How to Build a Mind

MAPS OF THE MIND

Steven Rose, General Editor

MAPS OF THE MIND

STEVEN ROSE, GENERAL EDITOR

—

Pain: The Science of Suffering
Patrick Wall

The Making of Intelligence
Ken Richardson

How Brains Make Up Their Minds
Walter J. Freeman

Sexing the Brain
Lesley Rogers

Intoxicating Minds: How Drugs Work
Ciaran Regan

How to Build a Mind

Toward Machines with Imagination

Igor Aleksander

Columbia University Press
New York

Columbia University Press
Publishers Since 1893
New York Chichester, West Sussex

Copyright © 2001 Igor Aleksander
All rights reserved

Library of Congress Cataloging-in-Publication Data
Aleksander, Igor.
How to build a mind : toward machines with imagination / Igor Aleksander.
p. cm. — (Maps of the mind)
Includes bibliographical references and index.
ISBN 0-231–12012–5 (cloth : alk. paper)
1. Artificial intelligence. 2. Imagination. I. Title. II. Series.

Q335 .A44225 2001
006.3—dc21 2001017431

Casebound editions of Columbia University Press books are printed on permanent and durable acid-free paper.

Printed in the United States of America

c 10 9 8 7 6 5 4 3 2 1

First published by Weidenfeld & Nicolson Ltd., London

Contents

PREFACE *vii*

1 Imagination and Consciousness *1*
2 Miletus: Where the Dreaming Begins *15*
3 Nineteen Fifty-eight: A Voyage Toward Interdisciplinarity *29*
4 The Ghost of Aristotle: An Influence Across Two Millennia *43*
5 Early Artificial Neurons and the Beginnings of Artificial Intelligence *57*
6 Liberating Philosophy: The Empiricists *71*
7 Canterbury: The First Machines *87*
8 Wittgenstein: A Brief Interlude *105*
9 The WISARD Years: Machines with No Mind *113*
10 Starting the Week with Consciousness *127*
11 MAGNUS in South Kensington and Pasadena *141*
12 On Being Conscious: The Ego in the Machine *161*

EPILOGUE *183*
FURTHER READING *189*
NOTES *191*
INDEX *197*

Preface

Doth glance from heaven to earth, from earth to heaven;
The lunatic, the lover and the poet
Are of imagination all compact.
One sees more devils than vast hell can hold—
That is, the madman. The lover, all as frantic,
Sees Helen's beauty in a brow of Egypt.
The poet's eye, in a fine frenzy rolling,
Doth glance from heaven to earth, from earth to heaven;
And, as imagination bodies forth
The forms of things unknown, the poet's pen
Turns them to shapes and gives to airy nothing

A local habitation and a name.
Such tricks hath strong imagination,
That if it would but apprehend some joy,
It comprehends some bringer of that joy;
Or, in the night, imagining some fear,
How easy is a bush supposed a bear!
—William Shakespeare, *A Midsummer Night's Dream*

In my laboratory we are doing work that is called "reverse engineering" of the human visual system. My senior research assistant Barry Dunmall and I are puzzled by the way in which fifty or so brain mod-

ules, each made of millions of cells (neurons), give rise to very accurate visual sensations of the world. At the same time, the brain can imagine previously seen things or things that have never been seen at all. We are staring at his latest model which shows up on the screen of a computer. In terms of twinkling patterns, it shows the activity of every cell in a simulated (scaled-down) chunk of brain that includes visual and language areas. Barry sets the thing off and we can see its eye roaming around a simplified scene made of colored, stylized fruit, and some of the neurons giving a very accurate representation of the scene, while others are trying to describe it.

IGOR: So you have now separated out the imagination net from the perceptual one.

BARRY: Yes. Last week we were talking about how the model was going wrong because if you asked it to imagine things, its perceptual area was overwhelmed by the imagining and it was hallucinating rather than imagining.

IGOR: This then supports the idea that in the brain imagination and perception are physically separate. The evidence from brain-scans is a bit ambiguous. We must bring this up the next time we have our joint meeting with the neuroscientists.

BARRY: What's not clear to me now is how both these things come into consciousness. They are in different parts. We're losing a grip on what's conscious and what isn't.

IGOR: I think it's okay. If we stick to Francis Crick's idea that a neuron that contributes to consciousness must fire according to world-centered events, this still happens. The imagination neurons are producing imagined world spaces. So if I think of my cat while I am looking at you, the vision of the cat has a place in my imagined field of view, but because of the separation it does not "feel" as if you and the cat get confused—it simply feels as if the two are somehow simply happening at the same time. The model's coming good now, let's stick with it and start looking at positional imagination clues like "think of an apple to the left of the banana."

I only recall this recent conversation because it shows that words like *consciousness* and *imagination* have crept into the world of computational machinery in much the same way as *knowledge*, *memory*, and *intelligence* have done in the past. I know that some will approve and others will be horrified at what engineers are doing by using words in this way. In fact, I believe that, while much care needs to be taken when coupling words such as "imagining" with machines, disapproval should not prejudice the exciting insights into mental phenomena that prompted me to write this book. Not everyone can join Barry and myself in our laboratory, but, through this book, I describe how, over many years, we have got to a point where working with imagining machines is beginning to tell us how real imagination and consciousness might come about.

The joy of being conscious, for me, lies in the amazing sense of freedom that I feel is available to my imagination. My consciousness not only puts at my disposal an accurate perceptual knowledge of the real world, but also enables me to imagine what I want to do in this world and allows me to make up worlds and actions that I may never have experienced. While most agree that the brain is responsible for our sensations, most will also say that it is now known how perceptual knowledge gives rise to imagination. My kind of engineering contains some simple, easily understood principles that are there solely for the purpose of explaining similar cause-and-effect circumstances in machines. To keep things simple, and to avoid the "oily rag" image that the word *engineering* evokes, I have described in this book both a biography of my own understanding, and some imagined encounters with philosophers who created the backbone for what needs to be understood.

The trouble with philosophical discussion is that in itself it has become very technical. It is often burdened by many "-isms." *Dualism*, *reductionism*, *epiphenomenalism*, and the like have become the currency that circulates through some texts and burdens the airwaves during public debates. My consciousness is a joy for me and not an "-ism." When I approach our machines in the laboratory, that's what I want to know: how does the mechanism of my brain translate into

the joy of my free imagination. The machine is just a tool that may help me in my quest. "Outrageous reductionism, scientific arrogance," someone once said to me. "You are ignorant of two and a half millennia of philosophy, you are flying in the face of what people wiser than you have concluded about consciousness."

I agree that it is very important to know what philosophers have said over this period of time. In my working life I took to reading about philosophy in parallel with the development of my engineering work, and I wish to reflect on this through this book. I believe that the likes of Thales of Miletus, Aristotle, Descartes, John Locke, Ludwig Wittgenstein, and many modern thinkers are driven by the same questions as I am. The difference is that we all have different tools and methods with which to produce some answers. I wish these thinkers to feature in this book not as austere intellectual figures, but rather as people who are using their own experience and the knowledge of their time to reflect on the nature of our mental lives. But how can my engineering experience with the machines of my time contribute to this? My late colleague Eric Laithwaite was very clear about this.[1] Many successes that engineers are proud of are present all around us as the products of natural evolution. Camouflage, the flight of insects, the vision of the night owl, are all more efficient than human technology at its highest. But the reason that even inadequate engineering is important is that it allows us to recognize and understand the working of complex machinery in nature and to use it as an inspiration for building useful machines.

And so it is with consciousness. This is a phenomenon possessed by living organisms which surpasses by far anything we can engineer in our laboratories. This is likely to be true for a long time, if not forever. But the process of developing progressively more competent neural machines has helped me personally to feel comfortable with how my brain might be causing my sensations. That which makes my consciousness different from that which can be built is an important part of the discussion. This difference holds some of the features of an explanation.

In addition to being an explanatory device, it is worth asking how such machine consciousness might also empower useful machines. The relationship between the artificial and real versions of consciousness remains just like that between the robot with vision and the night owl. The robot may be quite different, but understanding the properties shared by the two is sufficient to design robot vision systems inspired by the excellence of owl vision, and to understand owl vision better by knowing how robot vision can be designed. Not many would say that this diminishes our respect for the owl. So I hope that reading about artificial imagining machines built using human skill will encourage the reader to share in some of the excitement that Barry and I enjoy in playing with such machines.

Thanks

The fact that I have been able to write this book at all makes me grateful to more people than can be mentioned by name. At the time of writing I can list over sixty Ph.D. students without whom my own thoughts could not have developed at all. I also delight in getting the best family advice possible from a psychologist, a social anthropologist, and a biologist in the shape of Helen, Joe, and Sam. Their advice is not always on academic matters, but their fond tolerance for me and my bookish tendencies has been a joy for over a couple of decades.

How to Build a Mind

Imagination and Consciousness

Lights are Intelligences in our minds, whose force
We no more comprehend than here, in these
Glittering jewels, we can say how rose
Or sapphire blue or emerald steady shines,
Or what makes all the brilliant colours glow
Along the throat of the Arabian bird,
Whilst here, in milder air, her neck is grey
Or in the Polar void a brilliant white.
—A. S. Byatt, *Possession—Mummy Possest*

I imagine therefore . . .

As I sit tapping the keys on my laptop, I notice the oval table on which it is propped up with a newspaper so as not to wobble. The table is "rustic." It's the kind that costs too much in an antique shop, but finds its way to junk shops without too much difficulty. I look up and, through a large window, I see the Mont Scholastique—one of the many smoothly shaped tall hills of this part of the Languedoc. Pleasant memories of picnics at the ruined chapel at the top of the hill flood back. I look through a smaller window to my right and see the village of Octon across the cultivated valley of the Salagou River. Later today Helen and I will swim in the Salagou Lake . . .

Believe it or not, there is some purpose in this reverie. The reminiscence is made possible by an astonishing mixture of things: perception, memory of occurrences and feelings, plans and expectations. I am imagining things and through the use of words hoping that anyone who reads this can, using their imagination, share with me some of these experiences or, at least, believe them to be possible. But there's more: we are all capable of fantasy and of appreciating the fantasy of others. Great literature, poetry, science fiction, the movies, the stage are all part of what makes our mental world not just a dry topic called "mind" but a matter of supreme importance in our existence. Mind is the place in which we live. We can imagine the consequences of the actions of despots, appreciate the acts of benefactors, and live our lives according to principles. We live by our imagination, we continually add to our imagination, we trade our imaginings. We *are* imagining organisms. But are we just specific examples of imagining machines of which there are many? Do animals imagine? Could imagining machines be made? Can we understand what mechanisms in the brain bring this imagination about?

The usual word used to describe our active mental experience is *consciousness.* As words go, I feel that this one is a little tired. It appears in acrimonious debates in philosophy and science. It appears in the titles of many books, it is shunned by rigorous scientists and abused by not so rigorous ones. It has one significance for the anesthetist (will the patient jump off the table when stabbed with a lancet?) and a completely different and differing set of meanings for philosophers. Unashamedly, I want this book to be about consciousness, its wonders and pleasures. But I want to avoid the yawns and the pointless late-night conversations about its elusiveness. For this reason, I look for the force of consciousness in the power of imagination. I need to understand how my brain, an evolved machine of awesome complexity, can provide me with not only pleasurable reverie but also all the other elements of my mental life.

Imagination: Engineering and Philosophy?

Admitting that I am an engineer is (in the United Kingdom at any rate) a bit like standing up at a meeting of Alcoholics Anonymous and confessing to the error of one's ways. The trouble with the word is that it is associated with oily engines. In some parts of the world it is associated with the word *genius*—but I am far too humble to stress this. Nevertheless, being an engineer has been enormously helpful to me in understanding some of the principles that must be at work in the brain. In this book I want to pass these thoughts on without requiring anyone to know about engines of any kind, oily or otherwise. The point is that working with informational machines (some, but by no means all, of which are called computers) I find myself designing machines capable of handling the kind of stuff that makes up our imagination.

"Can a machine imagine?" is therefore a key question. If it can (and, clearly, I believe it can), how does its makeup distinguish it from one that cannot? The answer will not be revealed in the next paragraph or two but, hopefully, will begin to emerge by the end of the book. I hope that the reader will then join me in what is no more than a glimmer of an understanding. A glimmer may not be much to promise, but for me it has been a great step forward from understanding nothing. It is for this reason that the book needs to be read as a bit of a journey: a journey through my own past experience, and a journey through what some others have said about consciousness. The story will not be one of sifting systematically through scientific theory, as would be the case with a learned text, but a tale of pennies dropping in some kind of chronological order and in the context of my own developing comprehension. The backdrop does not stop with engineering, psychology, and neurobiology. A crucial context for such an understanding is philosophy. Philosophy is not something I learned through an engineering education but something I picked up as a fascinated external ob-

server. This too, therefore, provides resting stops in the proposed journey.

Machines I Have Known

Of course, I am not the first to have thought that an understanding of engineering can throw light on living organisms and their mental life. Many with similar aims inhabit the history of the twentieth century. In the early 1940s, the Massachusetts Institute of Technology (MIT) mathematician Norbert Wiener observed with some effect that the laws of control and information engineering which are used to analyze the automatic control of airplanes, rockets, and industrial plants apply to living beings as well. This became known as cybernetics and had a major influence in engineering, psychology, and management sciences in the 1960s. Also in the 1940s, the physician Warren McCulloch and the logician Walter Pitts, again at MIT, started making electronic models of brain cells and wondering at what point anything one could call "thought" begins to show up in networks of these. In the 1950s, British engineers Colin Cherry and Donald Broadbent became puzzled by the ability of human beings to attend to some things and filter out others. They modeled the processes using the electronic circuits of the day, and their models of attention have become central to the work of cognitive psychologists up to the present day. But these were clearly just beginnings. In chapter 3, I describe grappling with these ideas and trying to work out how much more needed to be done—it was a lot.

After a spell in industry and the completion of a Ph.D., I became a lecturer at Queen Mary College in London. I started to play with the design of single neurons. McCulloch had expressed himself in terms of the electronics of radio: volume controls, amplifiers, and the like. It would be inconceivable to build anything of any size in this way. Something had to be done in what was rapidly becoming a digital microcircuit age. But this was also the time that saw the beginning of the fashion for Artificial Intelligence (AI). Within this en-

vironment, thoughts of engineering inspired by the brain did not find a sympathetic hearing. In chapter 5, I describe how the computing world was becoming engrossed in making systems that seemed to do intelligent things, but where the intelligence came from the brilliance of programmers and the power of machines. I thought that this told us very little about how our own intelligence comes about. This was not a popular view. But I found the contrast between programmed AI and the ability of the neural networks in our brains—neural nets which might be investigated using the newest silicon techniques—surprisingly revealing.

In the late 1960s I moved to the University of Kent, which was then one of the "new universities." In the spirit of doing new things in a new university, I began making machines of a "neural" kind using the earliest digital microchips. The timing was bad as this approach was being thoroughly trashed by the Artificial Intelligentsia at MIT. Marvin Minsky and Seymour Papert published an influential book that elegantly destroyed the ideas of McCulloch and others about learning in neural systems, in favor of programmed intelligence. In chapter 7, I describe how this led to a trip to MIT, where rather than being convinced that the AI way was the royal road, I became even more deeply entrenched in my view that this approach will never satisfy anybody's curiosity about the brain.

However, another important penny dropped at that time: in order to have imagination, a machine must have neurons that interact with each other in such a way as to sustain their own internal encoding of things or embryonic "thoughts." In other words, they must have inner states (that is, firing patterns of neurons) capable of rich representations. An encounter with Stuart Kauffman, then a graduate student at MIT and now a recognized expert on emergent order in chaotic systems, led to the idea of these interacting nets of neurons having interesting emergent properties of their own. They have a neat way of arriving at their significant inner states in very few steps. This is pleasing because it correlates with most of our experience. Also, the contrary experience of, say, not

being able to put a name to a face, has an explanation as a property that emerges under some conditions in a neural net. These are properties that feature strongly in imagining nonliving nets, and possibly their living counterparts. In fact, these internal state structures and their emergent properties are the first major features that distinguish machines that can imagine from those that cannot. Easily said—but understanding and controlling these properties became not only a difficult research task, but one that required a great deal of concentration and led to turbulent times in my personal life at Kent.

A further move in 1974 took me to a chair at Brunel University. Both pressure from funding agencies and the desire to do something useful led to the design of WISARD, an early neural pattern-recognition machine. In chapter 9 we shall see how this machine, using neural networks, could recognize patterns as complex as faces but, alas, had no imagination at all. This allowed another important principle to fall into place. People talk far too glibly about "recognizing" things and then build machines that simply label patterns. There is a vast difference between recognizing patterns by labeling them correctly and knowing the objects that are perceived. Such knowledge is a happy resonance between imagination and perception, possessed neither by WISARD nor by the many neural pattern-recognition machines built over the last fifteen or so years. Something extra is required: yes, inner states are necessary, but they cannot be just any old inner states.

Working with neural systems received a boost in the 1980s. It became fashionable in the United States again with the work of innovators such as John Hopfield at the California Institute of Technology who, in 1982, published an elegant theory of inner states in nets based on principles of energy. So my desire to do something about machines with inner states capable of imagination returned just as I made another move: to Imperial College in London. In chapter 11, I shall outline how MAGNUS, a machine driven by inner states, was conceived, how it has developed to the present day,