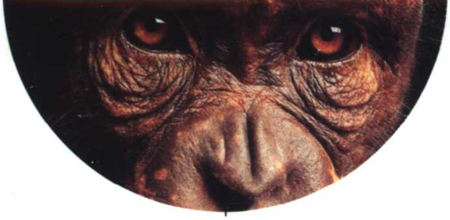
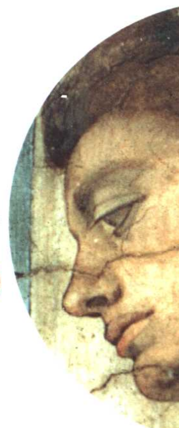


A SCIENTIST EXPLORES THE UNPARALLELED INTELLIGENCE OF THE HUMAN MIND



Are We Unique?

JAMES TREFIL

Are We Unique?

A Scientist Explores the Unparalleled Intelligence of the Human Mind

James Trefil

John Wiley & Sons, Inc.

New York • Chichester • Weinheim • Brisbane • Singapore • Toronto

This text is printed on acid-free paper.

Published by John Wiley & Sons, Inc.

Library of Congress Cataloging-in-Publication Data

Trefil, James.
Are we unique? : a scientist explores the unparalleled intelligence of the human mind / James Trefil.
p. cm.
Includes bibliographical references and index.
ISBN 0-471-15536-5 (cloth : alk. paper)
1. Human information processing. 2. Thought and thinking. 3. Intellect. 4. Artificial intelligence. 5. Psychology, Comparative. I. Title.
BF444.T74 1997
153—dc20 96-34338

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Also by James Trefil

A Scientist in the City

1001 Things Everyone Should Know About Science

Science Matters
(with Robert Hazen)

Reading the Mind of God

The Dictionary of Cultural Literacy
(with E. D. Hirsch and Joseph Kett)

Space Time Infinity

Dark Side of the Universe

Meditations at Sunset

Meditations at 10,000 Feet

A Scientist at the Seashore

The Moment of Creation

The Unexpected Vista

Are We Alone?
(with Robert T. Rood)

From Atoms to Quarks

Living in Space

To Harold Morowitz and the Monday afternoon gang at the Krasnow Institute, who never let me get away with anything.

Preface

There was nothing particularly unusual about the setting—it was just one more large, white room in a modern building full of large, white rooms. All around me computers hummed as young men and women stared intently at display screens. I was staring at one myself, one I had been directed to by my host, Ryszard Michalski. A slender man with long blond hair and engaging Continental manners, Michalski is recognized as a world leader in a new field of science that bears the innocuous name of "machine learning." His purpose that day was to introduce me to some of the products of his research, beginning with the simple little computer game I was playing.

Here's how it worked: The computer would display a set of figures on the screen. Each figure could have different body or head shapes—circle, square, triangle, etc.—in any of a number of colors. The figure could have a hat or not, carry a flag or not, and so on, and each of these appendages could also be of a different color. Obviously, there were a large number of possible figures, and the screen would show only twenty or so at a time.

With the figures on the screen, a voice came from the machine asking you to make up a rule about what figures would be "in" and which "out." For example, you might decide that only figures with square heads would be "in," or only figures carrying yellow flags. The machine then asked you to tell it a few of the figures that were in and a few that were out. It would then look at the information you gave it and try to figure out what your rule was.

A word about the voice. It was clearly not human, but in a way that was hard to pin down. It wasn't the metallic voice we've become accustomed to associating with robots in movies, but something at once more alien and more difficult to describe. It had a very strange pronunciation. For example, its version of "rule" came out something like "reeool." Not only the accent made the voice sound strange, however—many people speak English with accents more difficult to understand. There was also something about the voice that was different, as if the designers were determined to force you to acknowledge that you were being addressed by a machine, not by a human being.

After you had entered a few "ins" and "outs," the machine would be silent for a moment, then announce, in that uncanny voice, that it had discovered the "reeool." It would then proceed to tell you how you had made your choices. Sometimes, if it didn't have enough information to make a choice, it would ask you to enter a few more "ins" or "outs."

The game was fun. The machine would often come up with a rule that worked for the data you had given it, but wasn't the rule you'd had in mind. If you pushed it, you could often get it to come up with three or four alternate rules for the same set of data.

What was interesting about the machine's responses was that its programs seemed to imitate human reasoning. It seemed to be able to reproduce at some primitive level the ability to deal with ambiguity, with intuition, with all that ill-defined set of abilities that characterize human thought. Confronted with insufficient information, it made a guess. Only if the guesses didn't work did it ask for more information. Would a human behave any differently?

My first reaction was excitement. I was playing a game, of course, but the applications of this sort of system were obvious. As human beings delve deeper and deeper into the world, we use up the simple problems. Increasingly, the problems we want to solve are complex, and it's often difficult to see through the mass of data to the underlying simplicity that we believe is there. The trees make the forest all but invisible. A machine like this could cut through the impenetrable thicket of experimental results and give us a set of rules that might explain them. Once the rules have been called to our attention, of course, verifying and expanding them is relatively easy. It's discovering them that's hard.

In biochemistry, for example, there are a huge number of molecules that operate inside every cell. Are there simple rules that describe their structure and function? We believe so, although we've been able to discern precious few of them. Would a machine like this be able to help? And what about problems in ecology? There may be thousands of variables that describe a particular habitat. Which are important? Which should scientists consider when asked to make assessment of the impact of a new dam or factory? Often, we just don't know enough about the rules that govern the ecosystem to be able to say.

As the afternoon wore on and my probing of the machine's abilities became more subtle, other questions came to mind. "What if I fed all the information about Rembrandt paintings into this machine?" I asked. "Could it tell me how to produce a new one?"

Michalski smiled. "No," he said. "We really didn't know how to deal with the information in a painting."

I was reassured, but only momentarily. The machine I was using was the size of a suitcase—scarcely bigger than the PC on which I'm writing these words. What if someone gave these guys a CRAY? What if we waited ten years and gave them the best machine available then? What if you put a team of hot programmers on the problem for a decade? Would you then have a machine that would tell us how to make a Rembrandt, or worse yet, program another machine to do it?

Suddenly the room seemed a lot less cheery. It was still just as white, and the young people at their computers were still just as earnest and well-meaning. But what were they doing? Was I present at the creation of something that could someday make humans obsolete? Could these machines ever be "human," whatever that means?

I decided to put the machine to one more test. Acting quickly, at random, I fed it a set of "ins" and "outs" with no rule in mind. The machine sat and whirred for a long time, then announced that it had found the "reeool." "The rule is that the figure has either a square body or a yellow body, a hat or a yellow flag. . . ." The machine went on and on, spinning a rule of such complexity that I was flabbergasted. More important, I knew that no human being could ever have found that rule—that I was listening to the words of an intelligence that was totally alien.

As that eerie voice continued, I felt the layers of rationality and civilization that we all erect sliding away, giving way to the primitive fears that lie underneath. I could almost hear the wings of my ancestral vampires flapping in the last rays of a Carpathian sunset. I suddenly knew as certainly as I've ever known anything in my life that I was in the presence of . . . what? Evil? The word seemed both too strong and too insipid to describe what I was experiencing. Then it came to me. I was in the presence of *sacrilege*. What was being done in this very ordinary white room constituted nothing less than an assault on the human soul.

Strengthened by years of training, my rational mind quickly regained control. This was, after all, the last decade of the twentieth century, not some grade B movie. My host was no Victor Frankenstein, and those earnest young men and women weren't assistants named Igor. Great good would undoubtedly flow from their work—perhaps new cures for cancer, new tools in our constant struggle to feed mankind and conquer disease. I spent a few moments chatting with them, discussing the possibility of using their programs on one of my research problems. We arranged for me to give a seminar to explain the problem to the group, and I left.

Outside, in the vinyl-tiled hallway, I paused. I am not a religious man—it has been years since I've been to church. But I want to tell you, my friends, that whether it was rational or not, before I left that place I made the sign of the cross.

About the Author

Coauthor of the bestselling book *The Dictionary of Cultural Literacy* and the highly acclaimed *Science Matters*, and author of *1001 Things Everyone Should Know About Science*, James Trefil has published more than twenty books on science. A former Guggenheim fellow and a regular guest on National Public Radio, he is the Robinson Professor of Physics at George Mason University and a contributing editor of *Smithsonian* magazine. He lives in Annandale, Virginia.

Contents

1

Is There Anything Left for Us?

What a piece of work is man!
How noble in reason! . . .
In apprehension, how like a god!

—SHAKESPEARE, *HAMLET*, ACT 2, SCENE 2

Are human beings in some way different, unique in creation and in the eyes of God? Are we, in other words, special?

This is a very old question, and one that at first glance seems to have an obvious answer. Imagine, for example, that you were an extraterrestrial in a flying saucer approaching the planet Earth for the first time. Your instruments would pick up the usual signals of life—water vapor, oxygen, and so on. You'd settle in expecting to find a living planet with a complex ecosystem and then—whoa!—you'd see something absolutely startling. On this particular planet one species completely dominates the ecosystem. It's found virtually everywhere, and its works are of sufficient magnitude that they affect the rest of the planet's systems. By building large-scale reservoirs and lakes, for example, this species has actually slowed down the planet's rotation! It produces works of science and art beyond the capabilities of any other life form. If you knew anything about evolution and natural selection, you'd have to say "This is amazing! Something has happened here. These animals have found a new way to win at the evolutionary game—something no other species on the planet has developed."

Here are some of the characteristics of human beings that a hypothetical extraterrestrial might remark upon: the human ability to transmit nongenetic information from generation to generation via written and spoken language, the ability to devise massive technological systems that produce effects comparable to those produced by natural systems on the planet, the ability to use culture (rather than genetic selection) as a tool in the battle for survival, the ability to develop and manipulate abstract information, resulting in systems such as the ones we call science and language. Depending on its intellectual bent, the extraterrestrial might assign even more weight to the great moral systems embodied in the world's social and religious codes, to the complex aesthetic systems behind the buildings, paintings, music, and literature that permeate human life. All of these might seem to our hypothetical visitor (and to most of us as well) to make a pretty clear case for human uniqueness.

But appearances can be deceiving. It has lately become fashionable among intellectuals to ignore ways in which humans are different from other living things and concentrate on the ways in which we are similar. This trend, I think, may be partly fueled by a misplaced sense of egalitarianism among academics, but it is also based on a lot of good, interesting, and new data. As we shall see in chapter 3, we are starting to learn a great deal about animal behavior. We are starting to find that abilities that used to be thought of as uniquely human—the use of tools, for example, or of language—can sometimes be seen at some level in other living things. As astronomer Carl Sagan and coauthor Ann Druyan put it in their book *Shadows of Forgotten Ancestors* (Random House, 1992):

> Philosophers and scientists confidently offer up traits said to be uniquely human, and apes casually knock them down—toppling the pretension that humans constitute some sort of biological aristocracy.

So one assault on the notion of human uniqueness comes from studies of nonhuman animals. Some of what you read on this subject tends to be overblown, amounting to a claim that because animals can do some things that were previously thought to be uniquely human, there is no difference between humans and animals. I will argue that there comes a point where differences in degree become marked enough to become differences in kind. There is, for example, a rather profound difference between the toolmaking involved in a chimp using a stick to gather termites and that involved in humans building a jet aircraft or a skyscraper.

The traditional response to the question of the difference between humans and animals, of course, was the assertion that only humans possess a soul. In essence, this has the effect of removing the question of human-animal differences from the realm of scientific inquiry, a step I would be extremely reluctant to take.

It is possible, however, to approach this question without giving up either human uniqueness or scientific inquiry. To introduce an analogy we'll be using throughout the book, marking the precise boundary between humans and the rest of the animal kingdom is something like marking the boundaries of a city by traveling out along different highways and noticing where the City Limit signs are. If we pick enough highways on which to travel, and if we carefully note where the countryside ends on each one, then when we connect the dots we'll have a good approximation to the boundary of the city. In the same way, if we consider certain types of ability (the "highways") and look at animal studies, we will be able to find a point for each at which we can say, "Animals can go this far, and beyond here only humans can perform." In the end, we'll have produced a map of those activities and areas that are uniquely human.

The problem until now has been that people have tried to deal with this issue with too broad a brush. The question of whether animals have language ability, after all, isn't one that can be answered "yes" or "no." Instead, we should be asking about what level of language ability can be achieved by what animal under what circumstances, and use that information to "mark the city limits" in this area. When we are done with this process, we will be able to say precisely what separates humans from the rest of the animal kingdom, without necessarily being able to produce broad (and glib) generalities. And if those differences turn out to involve matters of degree rather than matters of kind, so be it. That's the nature of the world in which we live.

Actually, although the issue of animal intelligence is of interest to scientists and philosophers, I don't think most people are very concerned about the fact that some animals seem to have some limited ability to perform functions that most of us think of as uniquely human. Aside from some vociferous rear-guard action on the part of Creationists, most people (religious leaders included) made peace with the fact that human beings are part of the natural world not too long after Charles Darwin published *The Origin of Species*. We recognize that we are part of the great web of life that exists on this planet, and that means that we are related, both by blood and by descent, to every other part of the web. The reason this fact doesn't bother us is that we, like the hypothetical extraterrestrial, can see at a glance that no matter how close that relationship is, there is something *different* about us. And if intellectuals can't define that difference in precise language, who cares? To paraphrase former Supreme Court Justice Potter Stewart when he was pressed to give a definition of pornography, "We know it when we see it."

In fact, we know that this difference has mostly to do with the functioning of one human organ—the cerebral cortex in our brains. In chapter 2 we will explore the connection between *Homo sapiens* to the rest of the web of life and argue that from a biological point of view, it is this organ that provides the difference we seek—that pushes the "city limits" far away from us. Everything else about us, from our skeletons to the innermost working of our cells, is similar (and sometimes identical) to the ordinary run of things in the animal kingdom. As far as the human-animal boundary is concerned, we can rest assured that we are the same, yet different.

I should point out that the notion of human uniqueness is perfectly consistent with modern evolutionary biology. As we shall see in chapter 7, there are many species that have evolved unique adaptations over the millennia—think of the Venus's-flytrap and the bat's sonar navigational system, for example. Being unique doesn't necessarily make you special!

But as you probably guessed from the incident I recounted in the preface, my main concern does not lie with any imagined encroachment of the mental abilities of animals into the human sphere. With all respect to my colleagues in animal research, I don't see the day coming when a chimpanzee will be able to do a calculus problem or compose a symphony, no matter what kind of training it gets. Instead, I'm worried about a very new kind of incursion on traditional human space, one that comes from the machines that human beings, using their cerebral cortices, have built.

The reigning image we have of the human brain these days involves the machine we call the computer. In chapter 9 we will discuss a school of thought called "Strong Artificial Intelligence" (or Strong AI for short). The basic tenet of this school is that the brain is basically the same as a digital computer, although it's obviously more complicated than any computer we've built up to this point. If this is true, the argument goes, then it's just a matter of time before we are able to build a computer that's as complex and sophisticated as the human brain—just a matter of time before everything that our brains can do will be done by a machine. Although I'll argue later that this conclusion is far from obvious, it certainly defines another challenge to human uniqueness.

Go back to the city limit analogy. At any particular time, at any particular level of technology, we can define the boundary between human beings and machines by looking for the point beyond which machines cannot go. The idea would be to define a particular task ("make a painting," for example, or "solve this equation") and see how far a machine can go. On one side of the boundary, the machines can perform as well as (or better than) humans; on the other side, humans still hold sway, at least for the time being.

As we did when we were talking about the differences between humans and other animals, we can use this procedure to delineate a boundary between the domain of humans and the domain of computers. For the sake of argument, let's say the animal-human boundary marks the southern edge of our "city," the computer-human boundary the northern edge.

If the past few decades have seen a slow erosion of the notion that a wide gulf separates us from animals, they have seen a virtual disappearance of the notion that there is any gulf at all between the human brain and computers. You can see this in the widespread (and largely unexamined) assumption that the brain is just a complex computer. This idea takes its most extreme form in the notion that *Homo sapiens* is just a transitory form between the carbon-based life of the past and the silicon-based life of the future. Sometimes this hope leads to wild hyperbole, as when one artificial intelligence enthusiast a few years ago defined the goal of humanity to be "to build machines that will be proud of us."

If the computer jocks are right, if the brain is just a computer that we'll learn to duplicate and improve as time goes on, then the human-machine boundary can be expected to change quickly in the decades ahead. And this prospect, in turn, leads to an important and disturbing question: *When all the boundaries are*