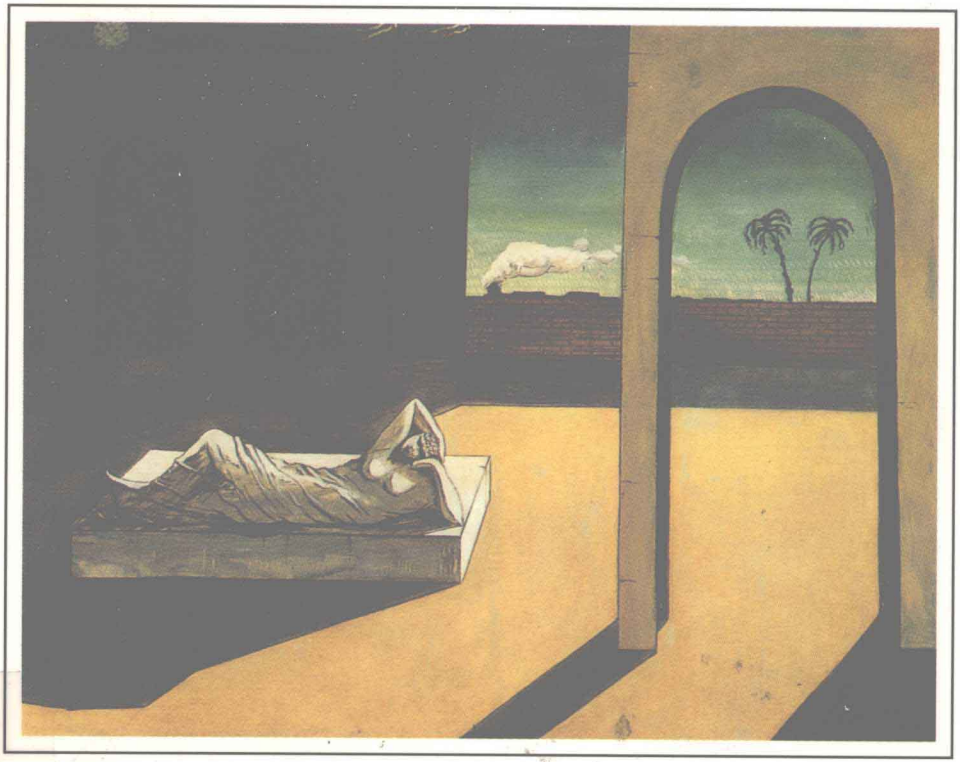
THE MUSE IN THE MACHINE

Computerizing the Poetry of Human Thought



David Gelernter

The MUSE in the MACHINE

Computerizing the Poetry of Human Thought

DAVID GELERNTER

The Free Press
A Division of Macmillan, Inc.
866 Third Avenue, New York, N. Y. 10022

Maxwell Macmillan Canada, Inc.
1200 Eglinton Avenue East
Suite 200
Don Mills, Ontario M3C 3N1

Macmillan, Inc. is part of the Maxwell Communication Group of Companies.

Printed in the United States of America

printing number

1 2 3 4 5 6 7 8 9 10

Library of Congress Cataloging-in-Publication Data

Gelernter, David Hillel.
The muse in the machine / David Gelernter.
p. cm.
ISBN 0–02–911602–3
1. Artificial intelligence. 2. Consciousness. 3. Cognition.
4. Emotions. I. Title.
Q335.G366 1994
006.3'01'9—dc20 93–49721
CIP

The MUSE in the MACHINE

THE FREE PRESS
A Division of Macmillan, Inc.
NEW YORK
Maxwell Macmillan Canada
TORONTO
Maxwell Macmillan International
NEW YORK OXFORD SINGAPORE SYDNEY

For my boys and my Jane

Acknowledgments

Acknowledgments are ordinarily a fairly cut-and-dried affair, but these are different. The summer of 1993, during which this book was completed, was the hardest of our lives. My wife and I needed lots of help, badly—and help was so generously forthcoming that I almost believe we will remember the summer more, in the end, for the kindness of family, friends, the community and total strangers than for the crime that occasioned it. If it in fact works out that way, isn't the only possible conclusion that good beats evil in the end?

Thanks are due first to our families—particularly to Cheri and Joel, to my parents and Jane's, Ed, Judy and other family members; and to our devoted dear friend Soledad Morales.

The manuscript would never have made it to the finish line without the extraordinary exertions of Nick Carriero and Chris Hatchell. I will always associate this book with their overwhelming generosity.

The Clemans, the Bialeks, Rabbi and Mrs. Benjamin Scolnic, the Agins, the Schwartzes, the Patkins and the Larrisons put themselves out repeatedly to help us with the children, which is the most important sort of help there is.

Among our friends and neighbors and my colleagues who helped

us in innumerable ways, large and small, the Lermans, Jane Milberg, Bob King, Susan Shapiro, Ed Keegan and Martin and Beverly Schultz were particularly indispensable. We deeply appreciate the help and good wishes of our friends Steve Nowick, the Arangos and the Jaggannathans, of Patricia MacDougall and the Yale Architecture Class of '83 (a truly great vintage!), of the Hamenachems, of the loads of people who brought us food, of the Yale police and particularly Officer Joe Vitale, of Ruth Anne Levinson and the JCC summer program staff, of Gail Brand and the whole Ezra community, Rabbi Weinberg and the B'nai Jacobites, President Levin, Provost Rodin and the Yale community *en masse*. I'm tempted to round out this huge list by appending the entire community of computer science researchers. Nothing moved and cheered me more last summer than the outpouring of good wishes that sloshed in daily from colleagues all over the world. People often write these notes because they feel they ought, without a sense that they will do much good; but they do immense good.

My graduate students carried on splendidly and cheered me up enormously. Thanks to all of them, and particularly Susanne Hupfer and Eric Freeman.

An academic is in bad shape without his books. Many thanks to Winston Atkins, Ellen Belcher and Paul Conway, whose expert ministration resuscitated so many of my tattered, burnt and drenched books and journals.

Last but (obviously) not least, my profound thanks to the medical people who so patiently reassembled me. They are a cast of thousands, too many to list. But the three whose exertions were most heroic are Rhonda Laidlaw, Marsha Dymarczyk and Dr. Henry Spinelli.

A few "normal" acknowledgments . . .

The research discussed here owes a great deal to my graduate students, but Scott Fertig in particular is responsible for much of the work and virtually all of the software that figures in this story. Dr. Mitch Sklar laid the basis years earlier with his own fine work.

My editor Susan Arellano did a great job with this quirky book. I am indebted to her, particularly for her patience.

Many parts of this book have gained from my conversations over the years with Professors Ivan Fox and Drew McDermott. (But

they are to be held wholly guiltless of any responsibility for the strange doctrines contained herein.)

Parts of Chapter 8 benefitted at an earlier stage from Barbara Harshav's superb editing and critical acumen. I'm only slightly ashamed to confess that if Mrs. Harshav liked this book and everyone else in the world hated it, I'd consider it a great success.

Every book I have written and every book I will write reflects, imperfectly, the wisdom of George McCorkle. He was an inspiring man.

And finally my wife Jane, from whom this summer I have learned the meaning of *eishet hayil. Ha-zorim b'dima, b'rina yiktsoru*.

Contents

Chapter One

A New Folk Psychology

It's hard to conceive offhand of a less promising consumer innovation than a computer that comes factory equipped with "emotions"—but here's a candidate: how about a "*spiritual*" computer? The spiritual computer spends its time pondering the mysteries of the universe, occasionally printing cryptic messages on its screen and otherwise ignoring the user altogether.

Here's what the "emotional" computer would do. You might describe a complicated medical case to it and ask for the diagnosis. The computer might give you a serious, telling answer, but add, ". . . still, I'm not happy with that; it doesn't feel right."

You might describe a complex legal case and ask for its advice. It's answered harder questions before, but on this occasion it might put you off with a comment about how the plaintiff reminds it of your sister.

You might describe an intricate stock deal and ask whether you should invest. In the past its advice has been solid—not infallible but better than any human's you've ever consulted, in part because it has billions of case histories down cold. But on this occasion it tells you, " 'Buzz off. I'm not in the mood. Let's talk about Jane Austen."

Who needs this kind of nonsense from a computer? Science does; in a broader sense we all do, because adding "emotions" to computers is the key to the biggest unsolved intellectual puzzle of our time: how thinking works. Oddly enough, our "emotional" computer will be capable of "spirituality" as well. No topic is further than spirituality from the interests of the researchers who are trying to understand the human mind. But as we will see, spirituality turns out to be central to cognitive psychology, and therefore to artificial intelligence, and therefore to computer science, and therefore to the whole history of science and technology.

Why these dramatic pronouncements and radical claims? Surely the science of mind is making fine progress without them.

Of course it is, but at the same time it suffers from a large and important problem. Here's an analogy.

Until 1759 sailors faced a big obstacle at sea: they only half knew where they were. Latitude was easily determined, but longitude was a mystery. Only when a meticulously accurate portable clock emerged at the end of a massive international research effort did the mystery of the longitude finally come clear (Landes 1983). To sailors, this development was a revelation; knowing half your position wasn't always a lot better than not knowing it at all.

Thought scientists today have a similar problem. The study of how thinking works is a big field—it encompasses philosophers of mind, cognitive psychologists, neurophysiologists and legions of frantically intense computer scientists bent on carrying off the greatest conjuring trick of all time, building minds out of computers. But thought science today is at sea. Despite monumental exertions, it has achieved a good grasp of no more than half the problem before it. Reasoning is one big part of human thought, and thought science has reasoning decently under control. Philosophers and psychologists understand it and computers, up to a point, can fake it. But there is one other big piece of the picture, which goes by many names: creativity, intuition, insight, metaphoric thinking, "holistic thinking"; all these tricks boil down at base to *drawing analogies*. Inventing a new analogy—hitching two thoughts together, sometimes two superficially *unrelated* thoughts—brings

about a new metaphor and, it is generally agreed, drives *creativity* as well. Studies (and intuition) suggest that creativity hinges on seeing an old problem in a new way, and this so-called "restructuring" process boils down at base to the discovery of new analogies. How analogical thinking works is the great unsolved problem, the unknowable longitude, of thought science. "It is striking that," as the philosopher Jerry Fodor remarks, "while everybody thinks analogical reasoning is an important ingredient in all sorts of cognitive achievements that we prize, nobody knows anything about how it works"—not even, Fodor adds (twisting the knife), in an "in the glass darkly sort of way" (1983, 107).

Thought scientists have exhausted themselves trying to solve the riddle. To cite only the most dramatic example, Roger Penrose (1989) posits by way of explanation an esoteric quantum mechanical freak of nature. Fodor (1983, 127) believes the answer is not merely unknown but unknowable—that such problems are just "bad candidates for scientific study."

I will present here a new model of human thought that puts analogy at the center of the action and offers what I claim is a direct, fairly simple explanation of how it works.

I'm about to tell a new and different story of human thought. But new and different does not mean concocted out of whole cloth. My method is more like the archeologist's who reconstructs an ancient vase out of its scattered shards. Most of my shards are important, intriguing, but (I claim) insufficiently appreciated results of experimental psychology. Some of my shards come from philosophy or computer science. Some—the most beautiful—are the work of poets, particularly the mind-obsessed English Romantics. A great poet has a more than average chance of possessing a mind that is, as William Wordsworth puts it, "haunted by herself." "In Egypt, Palestine, Greece and India the analysis of the mind had reached its noon and manhood," Samuel Taylor Coleridge writes in 1817, "while experimental research was still in its dawn and infancy" (Coleridge 1817/1975, 54). It's an overstatement, but an intriguing one. Modern thought science has no truck with poetry. But I am convinced that these poets can tell us deep, beautiful, *scientifically indispensable* facts about thought.

Like weathered pottery fragments, some of my pieces won't fit

exactly; others are missing and will require what appear (at first) to be big leaps of faith to reconstruct. But when everything is in place and a smooth, coherent shape emerges, those leaps will seem merely inevitable.

The result will be, in essence, a new "folk psychology"—cognitive science's slightly snide term for prescientific, commonsense psychology, the sort that can be done before even a single grant proposal gets funded. I will marshal a fair amount of scientific data to support my argument, but I will appeal more frequently to intuition and common sense. The results might or might not be convincing to scientists. I hope they are, but my main goal is to reach those readers for whom the human mind is not a profession but a passion. These aficionados are as likely to be poets, priests, gossips, or truck drivers as cognitive scientists. If I can show these people anything at all that broadens or deepens their grasp of this boundlessly fascinating, all-consuming topic even a little, I will be satisfied.

Dinner parties and boat rides

Here is my argument. Human thought is laid out in a continuous spectrum. Every human mind is a spectrum; every human mind possesses a broad continuous range of different ways in which to think. The way in which a person happens to be thinking at any given moment depends on a characteristic I'll call "mental focus." Focus can be high or low or medium; it changes throughout the day, not because the thinker consciously changes it, as he might consciously raise his arm, but in subliminal response to his physiological state as a whole. Fatigue (for example) makes focus go lower. Wide-awakeness makes it go higher.

Mental focus might sound like another way of saying "degree of alertness"; what's new is the way *cognition as a whole* changes in response to changing focus. High focus puts the thinker at the high end of the cognitive spectrum, and certain consequences follow. At the high end, thought is analytic and penetrating. It deals in abstractions and displays a "demythologized intelligence"—as the poet Robert Bly (1991) calls it—"that moves in a straight line made up of tiny bright links and is thereby dominated by linked facts." If a person's briefcase is stuck, he needs to open it and is in a high-

focus state, he will methodically run down the list of factors that cause briefcases to jam, plan a course of action and do it. His thoughts are well behaved. He has no doubt that they *are* (mere) thoughts; they do not impose themselves like hallucinations. Perceptions turn obediently into easily retrieved memories. In the future he will have no trouble recalling how he behaved during this particular stuck briefcase incident.

Almost all attempts to simulate thought on a computer have dealt exclusively with this narrow, high-focus band at the top of the spectrum.

As we set off down-spectrum, thinking becomes less penetrating and more diffuse, consciousness gradually "spreads out" and—this is a key point that I will spend much of the chapter explaining—emotion starts gradually to replace logical problem-solving as the glue of thought. The rest of the chapter pins down these vague pronouncements. When a briefcase jams and the owner's focus level is medium, instead of a cool logical analysis he is more likely to think "last time when I did this, it opened." Thought is less analytical, more concrete. He might simply give the thing a good whack. (Thought: a good whack usually helps out in situations like this.)

A few research efforts in artificial intelligence have attempted to reproduce this "medium-focus" type of thought—although, without noticing the spectrum itself.

Confronted with a stuck briefcase towards the bottom of the spectrum, the owner is likely not to solve the problem at all. He is more likely as he ponders the briefcase to find himself thinking "that was some hot muggy day when I bought this damned briefcase in Milford—did I overpay?—I pay more than I need to for most things. But I'm better than Bill Schwartz in that regard—Schwartz's dinner party last fall was sort of fun—Molly sure didn't want to go—she looked nice in that short midnight blue dress, though—Columbus Avenue, we got the thing in that shop around Seventy-sixth Street . . ."

Now suppose our thinker's focus is just a bit lower. His thought-stream might start off in the same way—might extend, say, through "Schwartz's dinner party last fall was sort of fun," and pause there for a while—various aspects of the party come to mind; and then, next thing he knows, the thinker might find himself contemplating

Long Island Sound as he crossed it on a ferry years ago, seated on the stern deck, admiring the glitter of the soft green water on a bright, hazy summer afternoon. That summer afternoon has no obvious connection of any kind with the Schwartz's dinner (which consisted, let's say, of an elaborate meal served on three end-to-end rickety card tables in a cramped SoHo apartment nowhere near Long Island Sound, and is memorable mainly for noise and the overwrought, grinning Schwartz's stories about the school board elections and Molly on her knees playing scrabble) but, for some reason or other, it comes to mind—

But this is a tremendously important phenomenon, this coming-to-mind seemingly out of thin air, when we are in a mental state that we might informally call "relaxed"—what I would call "low focus"—of thoughts bearing no evident relationship to their predecessors. Readers may doubt that it happens; just *because* it comes about in relaxed mental states only, the thinker rarely takes note of it, just passively experiences it. But psychologists have been aware of the phenomenon since at least 1823, when one Ludwig Börne (cited in Jones 1963, 160) wrote that, if you monitor your thinking uncritically for a few days, ". . . you will be amazed at what novel and startling thoughts have welled up in you." Modern studies that I will cite later back up this strange contention. Any reader who monitors his own thoughts for a while will discover the phenomenon for himself. Try it and you will see that it *does* happen.

Does it matter? Yes, enormously. These unexpected transitions from thought *A* to a seemingly unrelated thought *B* are (as I will discuss) exactly the occasions on which analogies are discovered and metaphors emerge. *A* and *B* don't *necessarily* make an analogy. The Schwartz affair and that trip across Long Island Sound might not be analogous in the least. But when analogies do emerge, they emerge in this way. It may be that, say, the Schwartz party and the ferry trip are two occasions on which our thinker had a wonderful time with a subtle undertone of anxiety, because of something that was going to happen the next day—

This mental leap from the noisy party to the placid boat ride is paradigmatic of the most significant unsolved problem of cognition. *Affect linking*, I will claim, is responsible for bringing these leaps about. They are not random (nor need they have anything to

do with repressed Freudian angst); they come about exactly when *two recollections engender the same emotion*, and they only happen towards the low-focus end of the spectrum.

Towards the lower end of the spectrum, affect linking causes creativity, metaphor, and in some cases spiritual mind-states to emerge. Other cognitive events accompany affect linking: thought grows ever more concrete. Recollection grows broader, more tangible and full of ambience and all-inclusive until eventually, a recollection becomes indistinguishable from a hallucination; and other things being equal, the illogic of dreaming waits at the bottom.

No computer program of which the author is aware attempts to simulate low-focus thought.

The cognitive spectrum provides us with a vantage point from which we can survey and make sense of human thought as a whole. More: it tells us something about the *dynamics* of human thought—its history over multiple time scales. Over the course of a day fatigue sets in and the character of thought changes. Over the first decades of life maturity comes about, and the character of thought changes. Over the millennia of human existence the modern mind gradually emerges, and the character of thought changes. I'll call these the Big Three cognitive transitions. They differ radically in character and take place over radically different time scales. But it's a curious fact that, if we view these three transitions from the spectrum's vantage point, they all three seem to tell the same underlying story. It is the story of gradual transition across the spectrum, from high focus towards low in the daily passage from awake to asleep; from low towards high in the development of a child or the emergence of the modern mind. The three transitions remain radically different, but the *underlying theme* turns out to be the same.

Now let me fill out this picture, make it more precise, and attempt to convince you that it is true.

The spectrum

Children have short attention spans.

There's nothing remarkable about that, but it perfectly epitomizes the *sort* of fact that appears to be wholly unconnected to top-